



OPEN ACCESS

Digital hermeneutics: scaled readings of online depression discourses

Inge van de Ven ,¹ Tom van Nuenen²

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/medhum-2020-012104>).

¹Department of Culture Studies, Tilburg University, Tilburg, Noord Brabant, The Netherlands
²Department of Informatics, King's College London, London, UK

Correspondence to

Dr Tom van Nuenen,
Department of Informatics,
King's College London, London,
UK; tom.van_nuenen@kcl.ac.uk

Accepted 21 June 2021

ABSTRACT

When it comes to understanding experiences of illness, humanities and social sciences research have traditionally reserved a prominent role for narrative. Yet, depression has characteristics that withstand the form of traditional narratives, such as a lack of desire and an impotence to act. How can a 'datafied' approach to online forms of depression writing pose a valuable addition to existing narrative approaches in health humanities? In this article, we analyse lay people's depression discourses online. Our approach, 'digital hermeneutics', is inspired by Gadamer's dialogical hermeneutics. It consists of a 'scaled reading' on five different scales: platform hermeneutics; contextual reading with term frequency—inverse document frequency (TF-IDF); distant reading with natural language processing topic modelling; hyper-reading with concordance views and close reading. Our corpus consisted of three data sets, from the blogs and message boards of, respectively, time-to-change.org.uk, a UK-based social organisation and movement that aims to counter mental health discrimination and alleviate social isolation by spreading awareness; Sane.org.uk, a leading UK mental health charity that seeks to help people in facing the challenges of mental illness and to improve quality of life; and the subreddit '/r/depression' on web discussion platform reddit. We found that the manner in which people express experiences of illness online is very much dependent on the specific affordances of platforms. We found degrees of 'narrativity' to be correlated to authorship and identity markers: the less 'anonymous' the writing, generally speaking, the more conventionally 'narrative' it was. Pseudonymity was related to more intimate and singular forms, with less pressure to conform to socially accepted and positive narratives of the 'restitution' type. We also found that interactive affordances of the platforms were used to a limited extent, nuancing assumptions about the polyvocality of online depression writing. We conclude by making a claim for increased cooperation between digital and medical humanities that might lead to a field of 'Digital Medical Humanities'.

When it comes to understanding experiences of depression, humanities and social sciences research have traditionally reserved a prominent role for narrative. Writers from such diverse fields as sociology, social psychology, literary studies, medical humanities, anthropology and philosophy all emphasise the pre-eminence and potential of stories in shaping the experience of, and discourse about illness. In trying to make sense of depression, and even as a potential pathway to recovery, narrativisation has demonstrable and demonstrated

merits. Writing and journaling have been proven beneficial as tools towards the recovery from traumatic and depressive experiences.¹ Stories help retroactively make sense of a depressive episode by placing it back in a diachronic continuity. They lend the 'self' a renewed sense of coherence.

Yet, experiences of depression are marked by elements that seem to resist narrative in a traditional sense, as a cause-and-effect trajectory with a beginning, middle and end.² As Peter Brooks has argued, the motor of narrative is desire: '[t]otalizing, building ever-larger units of meaning, the ultimate determinants of meaning lie *at the end*, and narrative desire is ultimately, inexorably, desire *for the end*'.³ Illness, by contrast, often manifests a lack of desire.⁴ Depression in particular is often discussed in terms of lacking desire and an impotence to act.

So where the telling of depression can help the subject re-integrate lived experiences into a continuity and coherent identity, the question is what kinds of narrative and non-narrative expressions offer the most accurate form in which to communicate the actual experience. Susan Sontag warned that '[n]o 'we' should be taken for granted when the subject is looking at other people's pain'.⁵ When it comes to representations of depressive experiences, understanding does not enter the reader's or viewer's head via osmosis. What forms of expression are available to us?

In this article, we examine online depression discourses, and ask if and how a data-centred approach to online forms of depression writing can pose a valuable addition to existing narrative approaches in health humanities. Our approach, digital hermeneutics,⁶ involves an oscillation between close and distant reading, as well as the familiar and the unfamiliar. Inspired by the dialogical hermeneutics of Hans-Georg Gadamer,⁷ digital hermeneutics allows us to use computational methods with sensitivity to ambiguities and subjective viewpoints. It helps confront and understand materials that reflect epistemological or experiential outlooks that might differ from our own. Applied to a corpus of online discourses relating experiences of depression, we expect this approach will yield insight into the ways in which depression is discussed and lived beyond, or in excess to, the existing narrative frameworks.

Our aim in this article is explorative. Rather than pursuing one angle exhaustively, we aim to show a range of possibilities for the analysis of an online corpus of depression discourse that such a 'scaled reading' opens up. Some of these warrant further, more extensive and systematic exploration. With



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: van de Ven I, van Nuenen T. *Med Humanit* Epub ahead of print: [please include Day Month Year]. doi:10.1136/medhum-2020-012104

this article, we seek to offer an example of how Medical Humanities research and Digital Humanities methodologies can inform each other. It can be seen as a preliminary step towards conceptualising a field of ‘Digital Medical Humanities’.

NARRATIVE TYPOLOGIES AND ALTERNATIVE AUTOPATHOGRAPHIES

Depression is an urgent public health problem that approximately three hundred million people will be confronted with in their life span according to statistics of the WHO. There is much we still do not know about its nature and causes, and it has a ‘heterogeneous’ clinical picture: the process and the response to treatment differ from patient to patient.⁸ Researchers in the social sciences and (medical) humanities have devoted considerable attention to lay people’s understandings of depression as an alternative and supplement to expert-driven biomedical research. Humanities research often moves past definitions of depression as illness, disease and pathology rooted in biomedicalisation, and looks at how we construct notions of illness, pathology and disease when it comes to mental health.

In many studies, the emphasis lies predominantly on narrative as a mode of expression. ‘Narrative’, here, roughly refers to all accounts that take the form of a story, that order events in time (with a beginning, middle and end structure), and involve changes over time that affect a protagonist. In telling the story of an illness, as part of the healing process, the ill person claims her body back from medicine and makes sense of the illness.⁹ Brooks goes as far as to claim that ‘*Mens sana in fabula sana*: mental health is a coherent life story, neurosis is faulty narrative’.¹⁰

One of the most influential typologies of illness narratives has been created by Arthur Frank in *The Wounded Storyteller* (1995). Frank distinguishes three categories: the restitution narrative, the quest narrative and the chaos narrative. In the restitution narrative, simply put, a healthy person first becomes sick and then recovers. Quest narratives, which occur when the illness initiates a spiritual journey, ‘meet suffering head on; (they) accept illness and seek to *use* it’.¹¹ The chaos narrative is most relevant for our purposes in this article as it is, in fact, an antinarrative, marked by absence of causality and closure: ‘Those who are truly *living* the chaos cannot tell in words. ... Lived chaos makes reflection, and consequently story-telling, impossible’.¹² Empirical studies of experiences of depression often resort to such narrative typologies.¹³ While such studies vary in the number of narratives or ‘tracks’ they employ, the narrative model as such is presupposed and remains unquestioned. We are dealing with not only a proliferation of depression narratives, but also a proliferation of *typologies* of depression narratives. Narrative conventions, however, are always both affording and limiting.

Narrative has its shortcomings when it comes to giving form to experiences of illness, as has been argued with respect to chronic illness and disability.¹⁴ This holds particularly true for depressive experiences¹⁵ which, due to cognitive impediments and distortions, are hard to describe accurately. Kiki Benzon describes depression as a disease whose ‘anarchic disconnections’, ‘bifurcation of mood’, and ‘murky distractedness’ make it incompatible with causal systems and linear teleology.¹⁶ Matthew Ratcliffe characterises depression as a loss of the possibility of meaningful events or projects.¹⁷

Of course, it should be noted that stories are never singular and exist in many different forms. They contain silences, partial speech, and ambivalent and uncertain speech, especially when people struggle to tell their story. Stories can also be complex and non-linear, full of gaps and contradictions and unspoken

aspects. They are shaped by available discourses. People struggle not just with depression but with giving it story-form, given the inadequacy of dominant available discourse. The dominant frameworks available to story one’s experience are often very limited and unable to address the complexity of them. This has been described by McKenzie-Mohr and Lafrance as ‘tightrope talk’.¹⁸ The term describes how people attempt to make meaning of their experiences, in a constant negotiation of agency and blame that is not possible in dominant narratives. Tom Strong in this respect writes of ‘discursive capture’, explaining how patients can lock themselves into a certain, singular discourse of understanding their experiences.¹⁹ Being captured, you are limited in the frames you have at your disposal for responding and understanding. Rather than using language, you become used *by* it. Miranda Fricker argues that epistemic injustice stems from ‘a gap in collective hermeneutical resources’²⁰ with respect to certain experiences, which make it hard for people to understand their own experiences. Power relations permit and constrain the generation of certain forms of knowledge.

As story-telling is a meaning-making action par excellence, it can be seen as ‘cure’ and impossibility at the same time: making a story out of depressive experiences often happens in retrospect, in remission. We are interested in different narrative and non-narrative forms of online depression discourses. Such discourses are abundantly available through online forums and blogging websites. Our focus in what follows is not in the first place on self-expression as a tool for recovery: rather, we explore the hermeneutic enterprise of trying to understand such experiences as an outsider. How can experiences of depression be communicated, and how can digital medical humanities research engage with these communicative acts? And what would ‘datafied’ forms of autopathography look like, or rather: how can data and stories be combined in such an approach?

Depression is cyclical,²¹ it often recurs in ‘episodes’, and it does not always have a definitive conclusion or end. We now shortly address two interrelated forms or representational strategies that offer an alternative to narrative models: the episodic structure and the database as a cultural or symbolic form. Both lead to a contemporary prevalence of paratactic, enumerative and list-like structures that emphasise the quantitative aspect of representation rather than stories with a beginning, middle and end.

Philosopher Galen Strawson (2004) argues against the normativity and universality of the notion that humans experience their lives as a story (the ‘narrativity thesis’). He distinguishes ‘episodic’ from ‘diachronic’ people. ‘Episodics’ conceive of their life as a series of discrete events, and the self as different people at the time of each event. Unlike ‘diachronics’, people that see their life as a single narrative with their self as the unchanging protagonist, ‘[e]pisodics are likely to have no particular tendency to see their life in Narrative terms’.²² They do not experience a continuous self that was there in the past, is the same person in the present, and will be there in the future. Strawson’s theory of episodic personalities has been used to nuance and complement the role of narrative in medical humanities research in general and depression in particular.²³ As experiences of depression are marked by a problematic relation to futurity, the episodic is a fitting form. It shows resemblances to Frank’s chaos narrative, which is typically expressed in a syntactic structure of ‘and then and then and then’, in a staccato pacing of words, alternated by ‘untellable silence’.²⁴

We argue that the episodic life experience becomes relevant in connection to recent changes in commonly used media ontologies, and what media theorist Lev Manovich (2001) has called

the database as a symbolic form. The cultural significance of the database—the format that underpins the computational interactions we engage in—has markedly increased during the last decades. This shift to an all-inclusive scope has consequences for the representational strategies we employ—in literature,²⁵ but also in culture at large. In the computer age, as Manovich argued, the database replaces narrative as our primary means of meaning-making. As a cultural form, the database marks a ‘new way to structure our experience of ourselves and of the world’.²⁶ Other than the narrative plot, the structure of the database is characterised by non-causality and lack of closure, and by the potentially endless addition of elements. As a result, we see more representations that organise data as a mutable, multilinear process. As a symbolic form, the database offers alternatives to the principle of causality that underlies the typical question ‘why me?’ and the quest to lay bare a chain of events it catalyses. These are supplemented by seriality or paratactic and open-ended, list-like structures that simply record one experience after another. In our analyses in this article, we focus on possible occurrences of such alternative structuring principles as mentioned here: the episodic and paratactic or serial, the ‘chaos narrative’ and database structures.

Before outlining the different levels of analysis and the platforms used for this research, a note on the relevance of *online* writing as an outlet for depression discourses is in order. The online trace data we are making use of can have been edited and (self-) moderated in a number of ways. At the same time, there are reasons for viewing the self-writing on these websites as congruent with methods in clinical psychology such as ecological momentary assessment (EMA). This method permits the research participant to report on symptoms, affect and behaviour close in time to experience, sampling many events or time periods.²⁷ EMA includes the use of traditional diaries, as they focus on collecting data repeatedly, in close to real time, and in subjects’ natural environments. Depression, relatedly, is a common thread in diary studies.²⁸ Online expressions on blogs, in this context, share pivotal traits with the diary, including the real time nature of the writing, the natural environment in which it can be performed and the potential repetition of the writing. It is no coincidence that blogging, in its early days, was often explained as ‘online dairying’.

Further, according to Benzon, online forms and genres of writing and publishing allow us to uniquely convey ‘the cognitive shifts and fissures’ of depressive thought in a way that is not possible in print publications. The ‘multitudinous narrative of depression’, she argues, is more akin to an organic ‘cultural narrative’²⁹ where such fissures and flows of thought are expressed in the structural dynamics of the environment itself, which is more pliable and collaborative than the analogue medium: ‘[I]nes of thought span out rhizomatically, emulating the cognitive mechanics of depression itself, whose salient feature is impaired concentration, a tendency to shift erratically among logical strands and the contours of myriad emotional states’.³⁰ Such polyvocal expressions lend themselves well to accurate descriptions of experiences of depression.

For the present study, the focus lies on online forums, blog sites and mental health discussion boards. In what follows, we present some preliminary results of patterns of depression that can be mapped out in web-based writing. For digital and medical humanists to engage in some form of systematic investigation, we first need to order this massive resource of textual data somehow. Where do we start and stop our reading, especially considering the overwhelming amount of sprawling, narrative and non-narrative information online?

DIGITAL HERMENEUTICS

In past years, different authors have discussed the potential role of hermeneutics in reflecting on the role of (digital) technology in mediating between human subjects and the world.³¹ Our approach here follows Gerbaudo’s concept of ‘data hermeneutics’ as a methodological reply to the anti-interpretative ideology of contemporary ‘dataism’. Gerbaudo stresses the need to identify procedures to select samples from data sets, so that they can be analysed in more depth. This, he argues, should be based on two processes: *qualitative sampling* procedures to reduce the size of (social media) data sets, and the development of a *data close reading* in relation to individual narratives, dialogical motivations and social worldviews.³²

Building on these points, our approach to digital hermeneutics combines (reflection on) interpretation with the use of computational methods and tools. It is inspired by the dialogical hermeneutics of Gadamer’s *Truth and Method* (1960), which commences from an ‘insuperable difference between the interpreter and the author’.³³ Gadamer understood the interpretive enterprise as a dialogue, or productive conversation with the text. In contrast to the positivist Enlightenment tradition in which subjectivity has to be left ‘at the door’ when commencing the analysis, Gadamer urges us to understand the existentialist tenet that prejudices are a function of our deep involvement and convergence with the world—and that they are necessary for any productive interpretative act. The only way to draw our prejudices into view, he suggests, is by their provocation when a text addresses us in its strangeness or unintelligibility.³⁴ The value of the dialogical perspective lies in this attempt to understand the other’s perspective without wanting to reduce it to one’s own or vice versa, and counters the idea that there is only one truth or one explanation.

Whereas Gadamer’s main focus is on the *historical* gap between text and interpreter, we aim to demonstrate that his theory is just as vital for bridging epistemological or experiential differences in a contemporary context. Updating the hermeneutic circle for digital humanities, we analyse online corpora in a circular motion that vacillates between the big data (‘n=all’) perspective of the whole, and a close reading of the part or the sample. For this paper, we have applied digital hermeneutics on five levels or scales, which we describe below. Such a scaled reading allows us to discern patterns in large-scale textual corpora, while also zooming in on the linguistic nuances of depression.

In our analyses, we use methods from natural language processing (NLP) for the second and third scales. As a field, NLP uses computers to process text and to assist in the identification of meaningful subjects and associations. It can help us to infer discursive regularity, topics or sentiments from unstructured textual data.³⁵ We worked with Jupyter, an open-source web application that allows one to create and share Python notebooks that contain live code, equations, visualisations and narrative text. Here, we outline the five scales and related tools we used:

1. *Platform hermeneutics* entail an examination of the specific affordances of the respective platforms, and how they relate to specific modes of self-expression and anonymity.
2. *Contextual reading*. What is the contextual horizon against which we can understand the linguistics, particularity of the respective corpora? *Tool: term frequency - inverse document frequency (TF-IDF)*.
3. *Distant reading* offers insight into the most important themes and semantic fields for each corpus. *Tools: Topic modelling*.

4. *Hyper-reading*. Traces patterns of discursive particularities and themes back to original context in the corpus. *Tool: Concordance views*.
5. *Close reading*. How can we analyse inherent and internal tensions, conflicts and irony that we found through the previous methods? What are the stylistic characteristics of the different modes of autopathography? *Close reading of a 'telling case'*.

The overall oscillation, ranging from platform hermeneutics to distant reading to a close reading of individual posts, could lead to a new iteration of the cycle.

THE DATA

The data has been generated in November 2019 from three websites that clearly differ in size, aims and affordances: blogs about depression have been retrieved from time-to-change.org.uk and sane.org.uk, and posts and comments were taken from the subreddit 'r/depression' on discussion platform reddit. The data cover all posts on these three platforms that had been generated at that time (see below). While the corpora by no means constitute 'big data' (in the sense that they are not 'big' enough to cause problems for typical computational methods), they are large and unwieldy enough to warrant the use of such methods, and certainly too large for human reading.

Time-to-change.org.uk (henceforth TTC) is a UK-based social organisation and movement that aims to counter mental health discrimination and alleviate social isolation by spreading awareness: it 'exists to end the stigma and discrimination experienced by people with mental health problems' (TTC). TTC includes a database of personal stories. The website clearly states the aims of these posts. The campaign closed at the end of March 2021. At the moment of data collection (26 November 2019), there were 1448 blogs, running from March 2007 to November 2019. We scraped the website using the well-known Python package BeautifulSoup, and filtered for posts tagged 'depression', creating a corpus of 601 blog posts, 492 929 words.

Sane.org.uk is a leading UK mental health charity that seeks to help people in facing the challenges of mental illness and to improve the quality of life for those affected. SANE employs mental health professionals and trained volunteers who provide emotional support, guidance (eg, through SANEline, an out-of-hours specialist helpline) and information for patients, family, friends and caregivers. They provide platforms for peer-to-peer support, including an online Support Forum with blogs on experiences, thoughts and opinions related to mental health issues. At the moment of data collection, 1490 such blogs had been published on this forum, with a total of 581 535 words. The data scraped using Python's BeautifulSoup package includes posts starting from January 2011 up to November 2019.

The third, and by far largest, corpus of texts has been scraped (extracted) from reddit.³⁶ Reddit is a web platform for social news aggregation, web content rating and discussion. Its members can submit content such as text posts, pictures or direct links, all of which are organised by distinct message boards, curated by interest communities, 'subreddits'. The subreddit 'r/depression' had 569 000 members at the moment of writing. It exists since 1 January 2009. Its motto is 'because nobody should be alone in a dark place', and its description reads 'peer support for anyone struggling with depression, the mental illness'. There are three primary ways to contribute on reddit: post a submission, post a comment, or vote on a submission or comment. The corpus that was gathered, after removing empty fields and fields shorter than 100 characters, includes 706 231 posts and 325 1544 comments.

Data (starting from 2009) were collected through reddit's application programming interface (a system of tools and resources in an operating system, enabling developers to create software applications). This allows access to submission, comment and user data. We made use of the 'timesearch' package. The final corpus consists of 191, 249, 252 words.

Before proceeding, we ought to note the ethical implications of using this approach—especially with regards to potential privacy violations when dealing with sensitive topics such as mental health. Mining data may present research with ethical issues such as right of use of information, privacy, right of access and so on. The Association of Internet Researchers has produced three major reports to assist researchers in making ethical decisions in their research (Franzke *et al*, 2019).³⁷ With regards to informed consent, we recognise that we are unable to retrieve consent from all the users of the three platforms we engage with; however, several steps have been taken to protect the identity of individual subjects, such as the deletion of usernames and other identifiable information, and the anonymisation of sources where used in this paper. We did not violate any terms of service of the platforms we have gathered data on. With regards to data storage, as our data include sensitive subject matter, all data were stored locally by one of the researchers. The reddit data were checked for deletions, and the data set is planned to be deleted on 26 November 2021, 2 years after scraping.

SCALE 1: PLATFORM HERMENEUTICS

On the level of platform hermeneutics,³⁸ we examined the websites and online platforms that host the blogs and discussion forums. This is important as from the next level onwards, text will be handled as data, disembodied and decontextualised, separated from its original environment. Platforms are designed environments that distribute control. They might be monetised, copyrighted, and negotiate forms and levels of visibility and privacy. As we will see, this environment in itself offers indispensable information that would run the risk of getting 'lost in datafication'. Platform hermeneutics regards the architecture of these websites, and the modes of sociality afforded by them. At this level, we look at the specific affordances of the platforms and how they relate to specific modes of self-expression and word use.

Both SANE and TTC have relatively low affordances: it is possible to upload a blog post of varying length; for SANE these contain mostly text and occasionally hyperlinks; on TTC, it is also possible to upload a photograph, other image or video. Some blog posts include links to the personal website of the blogger, where you can read more of their stories. Both pages consist of a list of posts, to be read in succession (although the sheer quantity of blogs amassed over the years would make that time-consuming), read one when it is published if you are a regular visitor, or skim the stories and click links that interest you. On TTC, it is possible to search for specific blog posts by entering keywords in a search bar. Both websites make it possible to interact by leaving a comment; yet notably, this option is rarely used.

Reddit has relatively high affordances compared with the other two, especially regarding interaction. The platform allows for up-voting and down-voting of posts and comments. Each link and comment displays a number of points (score), which corresponds loosely to the number of up-votes minus the number of down-votes a given item has received. The score of the post determines its placement in a feed, which is hierarchical. This affordance lends the subreddit a meritocratic structure: users earn the right to visibility through 'karma' points, currency that marks

the valuation of an individual's contributions to the community. These reward quantity and popularity of content, indicating that the user is an active, contributing member. r/ depression stands out from the larger platform as it is only allowed to post text, not images or unaccompanied web links.

Such technological affordances in turn shape the *social* affordances of each website: 'the possibilities that technological changes afford for social relations and social structure'.³⁹ Here, we see something interesting happening, as in all three cases, the interactive affordances (commenting on the blog posts, up-voting and down-voting on reddit), which in theory would foster sociality, are used to a very limited extent. This is especially surprising for reddit, which has interaction as one of its most salient affordances. Compared with other subreddits,⁴⁰ r/depression's posts and comments have a low response rate. We discerned a low level of reddit's typical subcultural aspects, like 'flairs' (tags that can be added to usernames, specific for a certain subreddit) and karma scores. This could indicate that people suffering from depression are likely to be either writers or readers online, but less inclined to engage (in)directly into dialogue. This limited use of social affordances gives the platforms a relatively monological structure, which nuances claims about the assumed polyvocality of online depression writing.⁴¹

With regards to governance, policies and moderation, the blogs on TTC and SANE are moderated and edited. This means there is a delay in publication: a blogger submits a text, the editors read and edit it and then place it online. When writing about suicide or self-harm, descriptions cannot be triggering, and TTC features clear guidelines for writing about these topics. It also includes policies for commenting, and it can take up to 5 days for comments to appear under a blog. SANE's moderation and policies, along similar lines, are made explicit in a file with 'creative expressions guidelines'.

On a subreddit, users post directly in a thread, so the production of discourse is more immediate. Reddit does use moderators to regulate the flow of content and monitor discourse: this is a voluntary position for any user with an interest in curating a particular subreddit. Yet freedom of speech is generally highly valued on the platform. Steve Huffman, CEO, stated that '[t]hese days, I tend to say that we're a place for open and honest conversations—'open and honest' meaning authentic, meaning messy, meaning the best and worst and realest and weirdest parts of humanity'.⁴² For r/depression, however, there are some strict guidelines included, prohibiting for instance 'tough love' as well as '[g]eneral uplifting or 'it gets better' messages: Encouragement is *not* helpful unless it integrates real, personal understanding of the OP's feelings and situation' (r/depression). Nor is it allowed to 'diagnose others or advocate *for or against* specific treatments or self-help strategies'. It is encouraged to be empathic, non-judgmental and 'interested in understanding the other person's experience'.

Another distinguishing feature of each platform for expression is the writer's respective level of anonymity. On TTC, writers can choose whether to post under a pseudonym or their real first name, to de-stigmatise mental illness. Many choose the latter, and often even add a picture. On SANE, it is not allowed for authors to include any identifying details, as stated in the policies. Authors post under a nickname. Reddit is pseudonymous as well: interaction on subreddits typically features elements of play and candour that one might not associate with traditional social-networking spaces that enforce a 'onename/real name' policy.⁴³ Many subreddits, notably those on politics, are marked by an ironic and 'in-joke' style of posting, making it hard to determine intent. This means that for us as interpreters

and researchers, we should be cautious in taking a comment to bear testament to genuinely felt resonance with content. There is no way of conclusively determining whether users are 'authentically' suffering from depression.

SCALE 2: CONTEXTUAL READING

Keeping these platform characteristics in mind, we turn to the data sets. With a 'contextual reading', we compare the corpora in terms of topical content. First, we lemmatised the corpora, meaning that we removed inflectional endings and returned words to their base or dictionary form ('lemma'), and we filtered them for nouns. We then applied TF-IDF, a simple, often-used technique to trace significant terms. TF-IDF stands for 'term frequency—inverse document frequency'. It is used in text analysis to find differences in textual corpora.⁴⁴ Term frequency (TF) gives us the frequency of the word in each document in the corpus. It is the ratio of the number of times the word appears in a document compared with the total number of words in that document. It increases as the number of occurrences of that word within the document increases. Each document has its own tf. Inverse document frequency (IDF), then, is used to calculate the weight of rare words across all documents in the corpus. The words that occur rarely in the corpus have a high idf score. Combining these two, we come up with the TF-IDF score (w) for a word in a document in the corpus. This score allows us to measure how isolated or specialised each community is, by looking at the proportions of actual in-words. TF-IDF can help bring out formal particularities for each of our three data sets.

Below, you see the 20 most typical nouns for r/depression, as compared with the other two sets:

reddit 0.595
fuck 0.393
semester 0.281
apartment 0.176
edit 0.174
tl 0.17
fucking 0.137
subreddit 0.124
gf 0.11
weed 0.1
dont 0.096
gpa 0.094
bullshit 0.092
ssri 0.09
roommate 0.088
zoloft 0.071
highschool 0.07
wellbutrin 0.069
rent 0.068
sophomore 0.066

Note that these words do not necessarily pertain to depression: they are simply the most frequently occurring words that are specific for this particular corpus. From this list, we can derive information about the users. For instance, it becomes plausible that the average reddit user is relatively young, as suggested by demographic markers like 'roommate', 'highschool', 'sophomore', 'gpa' and 'semester'. Swear words and mention of weed (possibly as a form of self-medication) can be explained by the relatively loose style of reddit's moderation compared with the blog sites. SSRI (Selective serotonin reuptake inhibitors), Zoloft and Wellbutrin are types of antidepressant medications: reddit users apparently write more about particular brands of antidepressants than those on the other two websites.

The most distinctive words in time to change (TTC), by contrast, refer to the affordances of blogging. The top terms are names. These names are likely not those of the author, as the author field was not taken into consideration. The inclusion of other people as ‘characters’ might indicate a stronger inclination towards narrative than is the case with the reddit data; a hypothesis that is to be tested on the next levels or scales of analysis. Further, there is an indication that this data set includes words related to knowledge production regarding depression, such as ‘stigmatisation’, ‘anosognosia’, ‘reassessment’ and ‘deprivation’. The SANE data set, finally, contains similar cognitive terms with a high TF-IDF score such as ‘psychodiversity’ and ‘philosophy’, but also ‘donation’, referring to the associated charity.

SCALE 3: DISTANT READING

For the next scale, we engaged in the common NLP approach of topic modelling, which enables us to trace semantic patterns for each corpus. Topic modelling programmes automatically extract topics from texts, taking a single text or corpus and searching for patterns in the use of words. The topic model that was built made use of Latent Dirichlet Allocation in the widely used Scikit-learn package for Python.⁴⁵

The assumption behind this machine learning technique is that documents consist of multiple topics, which are considered as ‘hidden variables’ that reflect the thematic structure of a collection. A topic model is built without semantic assumptions on the part of the researcher: the technique is ‘unsupervised’ and finds relationships between words without knowing what these words mean.

We filtered our corpus for nouns, in order to focus on the topical content of our corpus. Next, we needed to determine how to decide the optimal number of topics. The number of topics that makes for an accurate model can be tested through the calculation of coherence scores (although it has to be acknowledged that this method does not always fit perfectly with human intuition).⁴⁶ The measure, called ‘c_v’, indicates the relative distance between words within a topic (ie, how often they appear together in documents). For SANE, we selected a model of 20 topics (coherence .46), and 10 topics for time-to-change (coherence .38). For the subreddit r/depression, the initial coherence score, on a topic model with 44 topics, was 0.40 (see figure 1). Note that these coherence scores are relatively low; higher scores might be attained by further preprocessing and/or modifying the size of the documents used.

Potentially interesting topics for the r/depression subreddit included one around self-medication and mood manipulations, and others thematically reflecting a focus on suicide, emotions

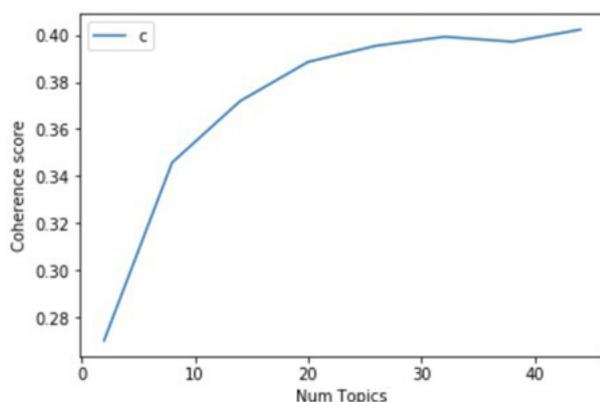


Figure 1 Coherence scores of topic model for r/depression.

and positivity.⁴⁷ Deemed most interesting for our purposes was a topic that mostly consists of temporal markers and daily routines. Topic 20, labelled ‘experience of time’:

day; sleep; night; bed; hour; wake; morning; tired; time; dream

We decided to focus on this topic for further explorations, as temporal indicators might point to alternative modes of representation that we could examine in addition to narrative models. To examine the temporal ‘rhythms’ of depression might offer insight in the way it is experienced as an incessant ‘now’ that forecloses futurity.

Examining the topic models for SANE, it struck us that several of the topics seemed to revolve around creativity, a theme that did not come out of the models for the other platforms.

Topic 6: *would; draw; colour; skin; smile; new; music; use; fire; poem*

Topic 7: *use; learn; emotion; present; emotional; poetry; therapy; skill; manage; art*

Topic 15: *song; dance; sing; lyric; mike; guitar; disagree; authority; diary; record*

This suggests that the support network of this community places emphasis on art and creative exercise as related to mental well-being and possible alleviation of distress. In SANE’s other topics, words describing moods and the weather recurred, giving rise to the hypothesis that the language used in these blogs is poetic, which fits in with the emphasis placed on creative expression.

The TTC blogs turned out to comprise too small a corpus to gather many relevant topics through topic modelling (with a coherence of only .38). We chose to focus on this one, as it has words indicating communication, like ‘tell’ (with the connotation ‘story’), ‘ask’, ‘share’ and ‘other’, as well as ‘struggle’:

topic 0: *back; tell; other; feeling; symptom; ask; struggle; medication; book; share*

Such words that refer to actions point to a more ‘narrative’ approach to expressions of depression.

The word ‘struggle’ in particular is salient. As a central structuring metaphor for depressive experiences, it could tell us something about the ways in which writers relate to their depression and the world beyond it. We take this up on the next scale of analysis.

SCALE 4: HYPER-READING

The fourth step is to read, in a more traditional fashion, sentences surrounding terms that we have found (‘local context’). We used Antconc, freeware corpus analysis toolkit for concordancing and text analysis.⁴⁸ Hyper-reading is an umbrella term for non-linear, screen-based and computer-assisted modes of reading, including search queries, skimming and scanning.⁴⁹ Hyper-reading allows us to intuitively and associatively trace our own interests and key across the data set, and thus helps us identify passages that contain a large number of meaningful features, and to select these for close reading. We engaged in a computer-assisted form of hyper-reading using concordances, allowing us to re-contextualise the significant words identified on the previous scale. Concordances are a tool for tracing key words across the corpus to their original lexical environments. Below is an example of such a concordance view for the term ‘struggle’ in a TTC data set (see figure 2). ‘Struggle’ as a trope automatically leads us to ask narratively ‘charged’ questions that helps us identify basic narrative patterns, such as: struggle *with* what, *for* what, *against*

119	and write what I'm about to. I struggle with depression. Am I depressed? I couldn't	timetochang
120	I'm Posted by russell84 26th Mar 2013 I struggle with depression and have done for a good	timetochang
121	have been a lot more open about my struggle with depression Depression can come and go and	timetochang
122	I thought I could trust about my 6 year struggle with depression. For years I have kept it	timetochang
123	difficult life situations. Even though I actively struggle with depression I can make a difference and	timetochang
124	and put it behind me. I continued to struggle with depression I continued to struggle with depr	timetochang
125	a lot more open and accepting of my struggle with depression in the last few years and	timetochang
126	Saturday/x92s for opening up about her struggle with depression. It is always hard for anyone	timetochang
127	positive one. After telling people of my own struggle with depression, others have talked to me about	timetochang
128	to struggle with depression I continued to struggle with depression, sometimes so severely I felt lik	timetochang
129	es. That everyone has mental health and to struggle with depressive thoughts on your own can just	timetochang
130	was concerned that if I told colleagues I struggle with general anxiety disorder and depression, it	timetochang
131	ography, where he talks candidly about his struggle with it. Driving to be the best, constantly	timetochang
132	feel better just saying out loud that I struggle with it". It's early days, but we	timetochang
133	r anxious. Many friends have revealed their struggle with low mood and anxiety Couple that with	timetochang
134	nds and acquaintances revealing their own struggle with low mood and anxiety. Many of us	timetochang
135	people who understood what it felt like to struggle with mental health issues I became really unwell.	timetochang
136	people who understood what it felt like to struggle with mental health issues. I met so many	timetochang
137	the people around me are unaware of my struggle with mental illness. After battling with my depre	timetochang
138	's a word that many of us who struggle with mental illness can relate to. This feeling	timetochang

Figure 2 Concordance view of TTC corpus in Antconc.

what? We can find preliminary answers to these questions by tracing the word struggle throughout the corpus.

Hyper-reading then yields sentences like: 'Everyday was a struggle and a battle against my own mind'; 'I struggle immensely with going to appointments, talking on the phone, being in crowds, going to the shops alone, meeting new people'; 'I remember the daily struggle I had trying to understand how and why I always felt so down'; 'a person struggling with a low mood and a lack of motivation won't be screaming at voices in their head and will instead struggle just to get out of bed in the morning'; 'I struggle mentally knowing that some of the people I am around don't know about my illnesses'; 'You internalise the struggle. You fight it alone'. In this manner, we can gain insight into the patterns of word use that people on TTC build around the notion of struggle, and hence how they live and experience their depression. The last quote prompted us to look for 'fight/ing' and that gives a very similar result, for example, 'fight against myself'; 'fight against stigma'; 'fight against mental health discrimination'; 'fight alongside a loved one'. Similar patterns could be mapped for 'battle': against 'a devastating illness'; 'myself' or 'my own mind'; 'stigma'. Thus, a semantic field emerged of militant metaphors.

Then we encountered the phrase 'fight your black dog' and decided to further explore that variety. 'Black dog' is a commonly used metaphor for depression. A user who actually created a puppet of their black dog, elucidates the benefits of thus concretising this disease: 'If I could build a big black dog puppet and give it life, I could maybe physically show depression instead of explaining how it felt. I could show depression as physically real'.⁵⁰ The image of the black dog has re-occurred from classical mythology through medieval folklore to modern times as a universal metaphor for depression and mental illness in general. Sir Winston Churchill famously used it to describe his dark moods. SANE even named one of their campaigns 'Black Dog': 'By encapsulating mental illness through the physical manifestation of the Black Dog, we enable people to visualise just how powerful, dominating and unpredictable it can be, while simultaneously affording them hope: dogs, like mental health conditions, can be tamed'.⁵¹ The varieties of the struggle or battle motif, along with the central black dog metaphor, make it very clear that the blogs on TTC follow a more conventional narrative pattern, which is foreshadowed in the organisation's name: 'time to change' already evokes a temporal transition or transformation that leads to a 'new chapter' in life.

A blog typically starts at the beginning of the author's mental problems, looking for causal patterns: 'Depression, my black dog, for me started early. I vividly remember breaking down, I must have been about 12, during a Saturday night family dinner due to the most intense feeling of futility and frustration with my life'. It includes opponents (often the mental illness or the writer herself/her own mind), friends, family and co-workers cast as helpers ('without my friends and family, I wouldn't have had the strength to persevere'; 'I have X in my corner'), and obstacles. Clark aligns such patterns with the Quest story:

[t]he quest or journey narrative organizes the experience of illness into a story of trials, helpers, ogres, or sorcerers (often psychiatrists offering meds), and a return as a subdued but wiser person. These shaping myths are just about impossible to escape.⁵²

As Sontag (1978) has argued with regards to tuberculosis, AIDS and cancer, the stories we tell are never completely our own. She urges us to 'detach' illness from these meanings: '[I]llness is *not* a metaphor, and ... the most truthful way of regarding illness—and the healthiest way of being ill—is one most purified of, most resistant to, metaphoric thinking'.⁵³ However, the 'battle mode' is hard to discard as its universality will warrant understanding. Narrative normalises and refamiliarises depression and makes it communicable.⁵⁴ TTC is the most traditionally 'narrative' of the three data sets, which can be explained by the mission of the organisation to destigmatise depression and to inspire people to seek or offer help, for which 'communicability' is obviously needed. Importantly, narrativity is also related to authorship and identity markers—we will come back to this in the conclusion.

For SANE, we decided to pursue the thread of the creative arts, selecting references to music for our hyper-reading. It appeared that music is associated with therapy, as these words often co-occur. Music also occurred in close affinity with meditation, art, poetry, beauty, dancing, scented candles, mindfulness, self-care and mental well-being, as well as adjectives like 'expressive' and 'good'. Other phrases include 'music for motivation' and 'music is a lifesaver', 'powerful tool' and a 'gateway' (into, respectively, laughter, self-care milestones and meditative practices). A number of blog posts inspire readers to curate their own personalised playlist for mood management. 'Poetry' occurs in similar ways: 'what really saved me was poetry'. This emphasis on individualised treatment, self-care and mood management brings out one of the goals of SANE as a charitable organisation:

'Increased capacity for self-management of mental illness and the process of recovery'.⁵⁵ Patients are increasingly encouraged in recent health policies to use self-help books, exercises and online tests in order to gain control over their mental health.⁵⁶ Lexical traces of the 'self-help' genre we found in the SANE data set are in alignment with Frank's 'restitution' narrative or the 'comic' storyline, which fits in with an emphasis on 'resilience' as a core ideal in neoliberal societies.⁵⁷

For the *r/depression* subreddit, we traced swear words that came out of the contextual reading with TF-IDF (term frequency - inverse document frequency), as they seemed indicative of the more immediate mechanics of publication on this platform and the loose moderation discussed at the level of platform analysis. Such linguistic particularities might indicate otherwise 'unfiltered' and relatively immediate expressions. Yet, tracing the lexical environments of words like 'shit' and 'fuck/ing' was not in itself the most illuminating angle to pursue, as these are extremely broadly applicable in many different contexts. 'Fuck/ing' especially can have both a negative ('fucking loser', 'depressed as fuck') or positive ('fucking awesome', 'fucking worth it') application. The use of 'fucking' as an adjective, however, does tell us something about the use of emphatic language to describe depressive experiences. Whether positive or negative, the word underlines that experiences are extreme. This aligns with a study that links experiences of depression with the use of 'absolutist' statements.⁵⁸

We then traced concordances of keywords for temporal demarcation that appeared in topic 20, labelled 'experiences of time', such as 'time' and 'sleep'. Like 'fuck', such words turned out to be too widely and variedly used to open up new angles. If there is something specific about the temporal experience of depression, it might become clear at a more structural level of announcement, and not necessarily in people literally using the word 'time'.

After these dead ends, we did a concordance search for 'I have' and 'I am', inspired by philosopher Jan Zwicky's notion that '[t]he experience of struggling with illness is the experience of the fundamentally metaphorical nature of self: one is, and is not, one's body'.⁵⁹ Is depression experienced as something you have, or something you identify with? Tracing 'I am' and 'I have' might lead to figures that give insight into the ways in which depressive experiences oscillate between subjects and object positions. Simile and metaphor (comparison) or synecdoche (identification) give potential insights into such subject/object relations.⁶⁰ Yet, rather than figurative language, our search for 'I am' yielded a large number of 'absolute' clauses like 'I am always' or 'I am never'. This could be explained by the fact that people suffering from anxiety and depression are considered to be more prone to thinking in absolutes.⁶¹ 'I have' revealed a dominance of the combinations 'I have been', 'I have experienced' and 'I have a feeling'.

SCALE 5: CLOSE READING

Hyper-reading led us to three 'telling' passages or samples with a high 'density' of the characteristic words that we here traced to their context in the data set—'I have a feeling' for reddit, 'music' (in combination with words like 'gateway' and 'meditation') for SANE, and 'struggle/battle' + 'black dog' for TTC. These were subjected to a close reading, which involved examining particularities of style as well as the role of emotion, ambiguity and irony.

Tracing words related to struggle and black dog in the TTC corpus led to the discovery of a post titled 'Advice isn't always

helpful when you're dealing with a mental health problem' posted on 4 April 2016 by user Brent (we anonymised the name). The blog is accompanied by a sizeable photograph of the author up close in a public space, a busy city street, smiling in an open way. This guided our reading attitude to some extent: if a person is brave and open enough to face the possible stigma of mental distress in order to create awareness, this warrants an attentive and empathic close reading.

The post aims to give advice to people seeking to offer support to a depressed loved one. It is thus directed at readers with no first-hand experience of depression. Brent gives a personal account of his depression, starting with his diagnosis and listing a number of consequences: the ending of a long-term relationship, the loss of jobs and friendships. Calling his depression his 'most defining feature', Brent explains how he had to step away from the notion it was something he could 'just shrug off and get over', which he compares to 'asking a terminal cancer patient to look on the bright side'. As this comparison physical terminal illness conveys, his depression is chronic. It is a battle: not just him against depression, but also a battle against ignorance. He uses the metaphor of a 'brick wall of negativity and lethargy' that prevents one from acting, and that one sometimes finds the strength to climb over. Friends and family support him, yet he also stresses that for those who have not experienced it, it is almost impossible to understand depression and anxiety. This leads to 'a sledgehammer made of well intentioned advice', suggesting activities like running, painting or meditating. In talking about his symptoms with friends, it is disparaging when they say things like 'that happens to all of us, mate'. Brent advises anyone who wants to help a person suffering from mental distress, to listen and refrain from offering advice. In the rare cases when people did this, 'a glorious silence descended on the room and my friend just sat there'.

In conclusion, the post makes a plea for silence and listening. Both this plea for silence and a stress on the incommunicability of depression point to the limits of language (which also explains the use of metaphors), and the value of simple co-presence and attention. Even though this post is 'narrative' in its adoption of the battle metaphor and emphasis on the role of others, this is clearly not a narrative of the 'restitution type'. It does not 'stitch up the wounds resulting from traumatic events or simply unexpected change',⁶² or normalise depression in any way by comparing it to experiences that might be familiar to the reader. It does not postulate an ending and does not try to give a positive or uplifting 'spin' to the story. Nor does it present depression after remission: the discourse is very much 'in medias res'.

For SANE, we pursued the combination of the word 'music' and words related to therapy and self-care, which led to a post by a user with the pseudonym RWTG, from 9 December 2014. RWTG appears to be a well-known member of the SANE community (or 'triber' as they are called) and an avid blogger. RWTG explains how his pseudonymous username was derived from Songs: Ohia's track 'I've been riding with the ghost'. They explain that despite the anonymity, blogging about their experiences is '*the most personal thing I have ever done*'. Hence, they chose an alias laden with personal significance. They call their alias 'the perfect alias, because it so succinctly summed up everything'.

The blogger interprets the 'ghost' in the song as a metaphor for depression, and then connects it to their own personal metaphor from their younger years, the 'corner man': 'The corner man is my ghost'. This figure, whom they have come to think about as personifying their depression, is on a journey with them. RWTG

analyses the lyrics to this and other songs and explains how these resonate with their personal experience:

There's something about that imagery, that juxtaposition of 'sirens' and 'silence', and the similarity of the words that I love. ... Sometimes it can feel like a siren – you've got so much crammed in to your head, voices are shouting, you feel like you're losing control of yourself and you can feel like the whole world is chaos. Then, it can be a silence – the cold emptiness, the loneliness of it all. It can be desolate, barren and bleak.

Thus, they describe the power of music, being evocative highly personalised in its meaning. They perform a similar personal analysis of The Smiths' songs 'Still III' and 'That Joke Isn't Funny Anymore', which at their worst moments made them realise they were not alone. Music is likened to a form of therapy more reliable than talking to friends: 'music is an eternal presence, consistently reliable and always dependable'. Like Brent, RWTG stresses the incommunicable nature of depression, but he also emphasises that many others undergo depression. Although the author chose to write under a pseudonym and did not include a personal photograph, after reading this post it feels like we know the person in an intimate manner, through their personalised close reading of the lyrics.

Tracing the combination 'I feel like I am' through concordances, we landed on a reddit thread consisting of a single post with one response. The post is an enumeration of a list of states prompted by the words 'I feel'. The list of feelings transmits a profound sense of loneliness and isolation that is hard to paraphrase; we decided to include it as an online supplemental appendix. Because of this enumerative structure of parataxis, listing one feeling after another without hierarchy or causality, the text is clearly the most non-narrative or even antinarrative of the three. The serial structure of affect here is on a par with Arthur's 'chaos narrative', as the post confronts us with 'time without sequence, telling without mediation, and speaking about oneself without being fully able to reflect on oneself'.⁶³ The chaos narrative knows no 'memorable past', nor a 'future worth anticipating', only an 'incessant present'.⁶⁴ There are no events, just feelings and states. There is no closure, no uplifting moment or resolution. The serial is made out of continuity and repetition ('I feel') with variation ('like a burden'; 'like a failure').⁶⁵ Because of its open-endedness, serialised writing complicates the attempt to present the self as a definitive whole. The self-understanding that emanates from such a mode of writing is episodic rather than narrative or diachronic, reflecting the instability of the depressed's sense of self over time.

The text strikes us as raw and authentic, immediate, even when using stylistic devices like repetition and contradictions or juxtapositions ('I feel like never eating again. I feel like eating everything in the world'). This immediacy stems from the repetition of 'I feel' in the present tense, and from the typos, spelling mistakes and emotional language. The author does not use humour or irony. The last sentence ('I'm sorry I just found your post, but If you need somebody to talk to I'll be here for you <3') was the only response to the post. Despite the relative lack of interactivity in this particular community, the presence of a supportive voice offering a listening ear speaks to the community's willingness to keep the conversation going, should the OP (original poster) want to.

CONCLUSION: COMING FULL CIRCLE

In this article, our aim was to show how hermeneutic, qualitative perspectives can be productively integrated with datafied

approaches. After a discussion of the role of narrative models in analysing depressive experiences in digital humanities research, we gave a brief overview of alternative forms of 'auto-pathography': the episodic, the database and online writing. Our approach of digital hermeneutics, inspired by the dialogical hermeneutics of Gadamer, was applied to a data set from three websites: time-to-change.org.uk, sane.org.uk and <https://www.reddit.com/r/depression/>. It consisted of analyses on five different 'scales' of text, gradually zooming in.

First, we engaged in platform hermeneutics, examining technical affordances in relation to specific modes of self-expression. The blog pages have relatively low affordances and low interactivity, where reddit has relatively high affordances, especially regarding interaction. In all three cases, the social affordances were used to a very limited extent and the platforms have a relatively monological structure. We also noted a difference in governance, with TTC and SANE being moderated and edited, causing a delay in communication, while reddit was more immediate with laxer policies of moderation. Another important difference lies in the pseudonymous nature of reddit and SANE versus the optional inclusion of personal identifiers on TTC.

Second, we performed a contextual reading, using TF-IDF to map linguistic particularities. Reddit's loose moderation style translated to colloquial language and profanities, as well as mentions of both prescription drugs and self-medication. Unsurprisingly, the blogs use a 'cleaner' language. Their data suggested a focus on knowledge production. In TTC, we also found mention of proper names, possibly indicating narrativity.

Third, we engaged in distant readings through the NLP approach of topic modelling. Potentially informative topics for r/depression included one consisting of temporal markers and daily routines, which might indicate non-narrative modes of representation. For SANE, we identified several topics revolving around creativity, indicating an emphasis in this community on art and creative exercise to restore mental well-being and alleviate distress. For TTC, we chose a topic that revolved around communication ('tell', 'ask', 'share', 'other', 'struggle'), expecting a focus on communicating and possibly narrativising depression.

On a fourth scale, we carried out hyper-readings of the corpora with concordance views, aiming to trace back salient words to their linguistic context of origin in the corpus. We traced back the word 'struggle' in the TTC data set, a revealing metaphor with regards to basic narrative relations (struggle *with* what, *for* what, *against* what?). This lay bare a semantic field of 'militant' metaphors. We also identified common usage of the old trope of the 'black dog' as personification for depression. Such tropes fit into narratives of the 'quest story' type, with adversaries, helpers and trials. For SANE, we selected references to music for hyper-reading. Music often occurred in close proximity to words semantically related to therapy as well as meditation and mindfulness, art, and poetry. This reflects an emphasis on individualised treatment, self-care and mood management in line with SANE's professed goals. This, we aligned to Frank's 'restitution' narrative, as well as an emphasis on 'resilience' as a core ideal in neoliberal societies. After a number of angles that gave us dead ends for the reddit data set, we did a concordance search for the combinations 'I have' and 'I am', where 'I have' yielded a dominance of the combinations 'I have been', 'I have experienced' and 'I have a feeling', which we chose to further explore.

Fifth and last, based on these findings, we engaged in a close reading of three 'telling cases'. We were attentive to particularities of style as well as the role of emotion, ambiguity and irony. The TTC post was the only non-anonymous and most 'narrative' in its account, yet rather than ending on an uplifting note

of restitution, it concluded with plea for silence and listening, co-presence and attention. The SANE sample, though pseudonymous, struck us as highly personal. The post exemplifies the emphasis on art as self-care we expected to find given the aims of SANE, yet in no way adheres to the promotion of 'resilience' associated with the restitution narrative. The post we selected for reddit was the most non-narrative or even antinarrative of the three, lacking events and closure, and reflecting an 'episodic' sense of self. The post affected us as raw and authentic, immediate and completely devoid of irony.

We found degrees of 'narrativity' to be correlated to authorship and identity markers: the less 'anonymous' the writing, generally speaking, the more 'narrative' it was (including helpers, opponents, widely known metaphors, battle motifs, etc). Interestingly, the pseudonymity of reddit and the relative 'immediacy' fostered by its moderation policies, here did not create an environment marked by the irony and play that reddit communities are typically associated with.⁶⁶ In case of 'stigmatised' topics such as mental distress, pseudonymous or anonymous writing has the potential to be more intimate and singular, with less pressure to conform to socially accepted and positive narratives of the 'restitution' type.

Another finding was that overall, the bodies of writing were less interactive, 'polyvocal' and participatory than expected. The interactive affordances were used to a limited extent. As suggested, this could indicate that people suffering from depression are more likely to be either writers or readers online, but are less inclined to engage (in)directly into dialogue.

Choosing to use social media platforms is agentic and active and a creative effort at coping. The conclusion that the sites were not very interactive is an interesting one. Many users may be seeking for an outlet and a way to voice their experiences without necessarily receiving judgement or advice. The platforms' relatively monological structure nuances statements on the polyvocality of online depression writing.⁶⁷

Our scaled readings proposed under the header of digital hermeneutics can offer a perspective on self-expressions of illness in addition to typologies of narratives. It aids researchers and medical humanities students in making use of a greater range of source materials to work with. These include narrative elements yet do not fit neatly into any preformulated typology. Compared with more traditional literary methods that pose a clear beginning and end point to their readings, digital hermeneutics, with its oscillation between close and distant, is better equipped to trace such non-narrative forms of life writing. It is important not to reduce pluriformity by generalising about 'online writing' on illness and mental distress. The merit of digital hermeneutics as an approach lies in its open and explorative nature, compared with narrative typologies.

Importantly, we demonstrated that the manner in which people express experiences of illness online is very much dependent on the specific affordances of platforms. Platform hermeneutics provides insight into how discourses of experiences of illness, as well as the forms of community built on them are shaped by platforms. We hope to have contributed to a growing body of scholarly work that analyses the impact of online media as fostering communities around illness. For instance, Sánchez Querubín (2020) analyses a range of social media uses as vehicles for 'illness storytelling'.⁶⁸ McCosker and Gerrard (2020) developed a practice-oriented research method based on hashtagging practices on Instagram.⁶⁹ Chateau (2020) analyses communities around depression memes.⁷⁰ Other studies provide insights into the narration and visualisation of a range of forms of mental distress.⁷¹ Further research could follow us in devoting attention to scale in data and stories.

The contribution of this article to medical humanities research lies less in offering new insights into depression as an illness, and much more in the methodological possibilities it opens up. This is not a matter of either/or, quantitative or qualitative, data or stories: they can productively be combined. To give insight in the interpretative (and sometimes intuitive) process of digital hermeneutics, we purposefully included speculations based on findings, hypotheses that led to dead ends, angles we later abandoned. Thus, we have been able to lay bare how depression online does not adhere to a numbers of narrative types, it rather consists of a plurality of forms. This pluriformity is already present in online writing and it can be better reflected in research if Medical Humanities research would foster collaboration with, and adopt methods from the Digital Humanities.

Twitter Inge van de Ven @IngevandeVen2

Acknowledgements A first version of this paper has been discussed at a meeting of PEERS (platform for discussion of research in progress among colleagues) in Tilburg University's department of culture studies, in November 2019. The authors thank participants of this seminar for providing constructive comments. A draft version has been presented as part of the workshop 'Data and Stories in Digital Healthcare: Mixed Methods for Medical Humanities' organised in name of the Volkswagen Stiftung in Trebbin, Germany in December 2019. The authors thank the organisers and participants of this event, especially Anita Wohlmann, for their insightful feedback.

Contributors IvdV has written the theoretical/literature review section and performed the close readings and hyper-readings of the selected passages, and written the conclusion. TvN has extracted the data from the online web forums and undertaken the methodological steps of the NLP methods under scales 1 and 2.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval Tilburg University's ethics board

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Inge van de Ven <http://orcid.org/0000-0001-7237-3779>

NOTES

1. James Pennebaker, 2004; Michael A Cohn, Matthias R Mehl, and James W Pennebaker, 2004
2. Jenny Slatman and Inge van de Ven, 2020
3. Peter Brooks, 1984
4. Arthur W Frank, 1995
5. Susan Sontag, 2004
6. Tom Van Nuenen and Inge van de Ven, 2020
7. Hans Georg Gadamer, 2004
8. A.J.W Van der Does and F.G. Zitman, 2008
9. Arthur W Frank, 1995
10. Peter Brooks, 1994

11. Arthur W Frank, 1995
12. Arthur W Frank, 1995
13. Ilka Kangas, 2001; Mervi Issakainen and Vilma Hänninen, 2016
14. Garden 2010; Anita Wohlmann and Madaline Harrison, 2019; Rebecca Garden, 2010a; Sara Wasson, 2018
15. Jenny Slatman and Inge van de Ven, 2020
16. Kiki Benzon, 2008
17. Matthew Ratcliffe, 2014
18. Suzanne McKenzie-Mohr and Michelle N Lafrance, 2011
19. Tom Strong, 2014
20. Miranda Fricker, 2007
21. Damien Ridge and Sue Ziebland, 2012
22. Galen Strawson, 2004
23. A. Woods, 2011; Jenny Slatman and Inge van de Ven, 2020
24. Arthur W Frank, 1995, 100
25. Inge Van de Ven, 2019
26. Lev Manovich, 2001
27. Shiffman, 3
28. Thiele, 5
29. Benzon, 146
30. Benzon, 148–49
31. Alberto Romele et al., 2020; Paolo Gerbaudo, 2016; Rafael Capurro, 2010; Tom Van Nuenen and Inge van de Ven, 2020
32. Paolo Gerbaudo, 2016
33. Hans Georg Gadamer, 2004
34. Hans Georg Gadamer, 2004
35. Carina Jacobi, Wouter van Atteveldt, and Kasper Welbers, 2016
36. Reddit is owned by the American company Advanced Publications. It needs to be noted that the other two websites studies are UK-based and that most of their users are as well; reddit has members from all over the world. This may be considered a limitation of the comparative approach, as national differences in healthcare policies and culture might influence how people discuss depression online.
37. A. S Franzke et al. (2020)
38. Tom Van Nuenen, 2018
39. Barry Wellman, 2001
40. See for instance interactivity on the subreddit '2meirl4meirl' which is devoted to memes about depression (Lucie Chateau, 2020).
41. Kiki Benzon, 2008
42. Andrew Marantz, 2018
43. Adrienne L Massanari (2015), *Participatory Culture, Community, and Play: Learning from Reddit* (New York Peter Lang, 2015).
44. Alternative methods here would be so-called feature-selection type statistical methods such as chi-squared and log-likelihood. Alternatively, AntConc has a corpus comparison and feature extraction function called 'keyword extraction'.
45. Fabian Pedregosa et al., 2011
46. Jonathan Chang et al., 2009
47. Topic 8, which we labelled 'self medication': "0.042*drink" + 0.035*drug" + 0.024*smoke" + 0.024*make" + 0.024*good" + 0.022*bad" + 0.022*feel" + 0.021*alcohol" + 0.020*problem" + 0.020*time"); Topic 34, which we labeled 'suicide': "0.069*suicide" + 0.061*kill" + 0.059*life" + 0.053*die" + 0.029*live" + 0.028*suicidal" + 0.026*end" + 0.026*death" + 0.025*thought" + 0.021*attempt"); Topic 14, which we labelled 'emotions': "0.490*feel" + 0.081*feeling" + 0.025*emotion" + 0.020*anymore" + 0.014*make" + 0.014*sad" + 0.013*empty" + 0.012*sadness" + 0.011*numb" + 0.010*tired"); Topic 32, which we labelled 'positivity': "0.180*happy" + 0.059*make" + 0.054*day" + 0.045*feel" + 0.029*happiness" + 0.028*birthday" + 0.024*smile" + 0.023*people" + 0.021*good" + 0.020*time").
48. <https://www.laurenceanthony.net/software/antconcl/>
49. James Sosnoski, 1999
50. For privacy reasons, we decided to not include titles and dates of, and hyperlinks to, particular posts.
51. http://www.sane.org.uk/what_we_do/black_dog
52. Hilary Clark, 2008
53. Susan Sontag, 1978
54. Deborah Flynn, 2010
55. http://www.sane.org.uk/what_we_do/aims_outcomes
56. Gaston Franssen and Stefan van Geelen, 2017
57. Brenda Dyer, 2008
58. Mohammed Al-Mosaiwi and Tom Johnstone, 2018
59. Jan Zwicky, 2008
60. Anita Wohlmann, 2018
61. Mohammed Al-Mosaiwi and Tom Johnstone, 2018
62. Hilary Clark, 2008
63. Sara Wasson, 2018
64. Sara Wasson, 2018
65. Anita Wohlmann and Madaline Harrison, 2019; Inge Van de Ven, 2016
66. Adrienne L Massanari (2015).
67. Kiki Benzon, 2008
68. Natalia Sánchez Querubín, 2020
69. Anthony McCosker and Ysabel Gerrard, 2021
70. Lucie Chateau, 2020
71. Megan A Moreno et al. (2016); Anthony McCosker (2017); Anthony McCosker 2018, Andrea LaMarre and Carla Rice (2017); R C Brown et al. (2018); Debbie Ging and Sarah Garvey (2018); Yukari Seko and Stephen P. Lewis (2018)

BIBLIOGRAPHY

- Al-Mosaiwi, Mohammed, and Tom Johnstone. "In an absolute state: elevated use of Absolutist words is a marker specific to anxiety, depression, and suicidal ideation." *Clinical Psychological Science* 6, no. 4 (2018): 529–42.
- Benzon, Kiki. "A Dark Web: Depression, Writing, and the Internet." In *Depression and Narrative: Telling the Dark*, edited by H Clark, 145–56. Albany, NY: SUNY, 2008.
- Brooks, Peter. *Reading for the Plot: Design and Interaction in Narrative*. New York, NY: Knopf, 1984.
- . *Psychoanalysis and Storytelling*. Oxford, UK: Blackwell, 1994.
- Brown, R C, T Fischer, A D Goldwich, F Keller, R Young, and P L Plener. "#cutting: Non-suicidal self-injury (NSSI) on Instagram." *Psychological Medicine* 48, no. 2 (2018): 337–46.
- Capurro, Rafael. "Digital hermeneutics: an outline." *AI & SOCIETY* 25, no. 1 (2010): 35–42.
- Chang, Jonathan, Jordan Boyd-Graber, Chong Wang, Sean Gerrish, and David M. Blei. "Reading tea leaves: how humans interpret topic models." *Advances in neural information processing systems* 32 (2009): 288–96.
- Chateau, Lucie. "Damn I Didn't Know Y'all Was Sad? I Thought It Was Just Memes": Irony, Memes and Risk in Internet Depression Culture." *M/C Journal* 23, no. 3 (2020).
- Clark, Hilary. "Introduction: Depression and Narrative." In *Depression and Narrative: Telling the Dark*, edited by H Clark, 1–14. Albany, NY: SUNY, 2008.
- Cohn, Michael A, Matthias R Mehl, and James W Pennebaker. "Linguistic markers of psychological change surrounding September 11, 2001." *Psychological Science* 15, no. 10 (2004): 687–93.
- Dyer, Brenda. "Winter Tales." In *Depression and Narrative: Telling the Dark*, edited by H Clark, 41–53. Albany, NY: SUNY, 2008.
- Flynn, Deborah. "Narratives of melancholy: a humanities approach to depression." *Medical Humanities* 36, no. 1 (2010): 36–9.
- Frank, Arthur W. *The Wounded Storyteller: Body, Illness, and Ethics*. Chicago & London: University of Chicago Press, 1995.
- Franssen, Gaston, and Stefan van Geelen. "Self-Management as management of the self: future directions for healthcare and the promotion of mental health." *Philosophy, Psychiatry, & Psychology* 24, no. 2 (2017): 179–84.
- Franzke, A. S, A. Bechmann, M. Zimmer, and Ess C. M, Association of Internet Researchers. "Internet research: ethical guidelines 3.0.", 2020. <https://aoir.org/reports/ethics3.pdf>. Accessed 16 Sep 2021.
- Fricker, Miranda. *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press: Oxford, 2007.
- Gadamer, Hans Georg. *Truth and Method*, 2nd ed. London & New York: Continuum, 2004.
- Garden, R.. "Disability and narrative: new directions for medicine and the medical humanities." *Medical Humanities* 36, no. 2 (2010): 70–4.
- Garden, Rebecca. "Telling stories about illness and disability: the limits and lessons of narrative." *Perspectives in Biology and Medicine* 53, no. 1 (2010a): 121–35.
- Gerbaudo, Paolo. "From data analytics to data Hermeneutics: the continuing relevance of interpretive approaches." *Digital Culture & Society* 2, no. 2 (2016).
- Ging, Debbie, and Sarah Garvey. "Written in these scars are the stories I can't explain": A content analysis of pro-ana and thinspiration image sharing on Instagram." *New Media & Society* 20, no. 3 (2018): 1181–200.
- Issakainen, Mervi, and Vilma Hänninen. "Young people's narratives of depression." *Journal of Youth Studies* 19, no. 2 (2016): 237–50.
- Jacobi, Carina, Wouter van Atteveldt, and Kasper Welbers. "Quantitative analysis of large amounts of journalistic texts using topic modelling." *Digital Journalism* 4, no. 1 (2016): 89–106.
- Kangas, Ilka. "Making sense of depression: perceptions of melancholia in lay narratives." *Health: An Interdisciplinary Journal for the Social Study of Health, Illness and Medicine* 5, no. 1 (2001): 76–92.
- LaMarre, Andrea, and Carla Rice. "Hashtag Recovery: #Eating Disorder Recovery on Instagram." *Social Sciences* 6, no. 3 (2017), 68.

- . "Engaging mental health online: Insights from *beyondblue*'s forum influencers." *New Media & Society* 20, no. 12 (2018): 4748–64.
- McCosker, Anthony. "Tagging Depression: Social Media and the Segmentation of Mental Health." In *Digital Media: Transformations in Human Communication, 2nd ed.*, edited by Paul Messaris, and Lee Humphreys, 31–9. New York: Peter Lang, 2017.
- McKenzie-Mohr, Suzanne, and Michelle N LaFrance. "Telling stories without the words: 'Tightrope talk' in women's accounts of coming to live well after rape or depression." *Feminism & Psychology* 21, no. 1 (2011): 49–73.
- Moreno, Megan A, Adrienne Ton, Ellen Selkie, and Yolanda Evans. "Secret Society 123: understanding the language of self-harm on Instagram." *Journal of Adolescent Health* 58, no. 1 (2016): 78–84.
- McCosker, Anthony, and Ysabel Gerrard. "Hashtagging depression on Instagram: towards a more inclusive mental health research methodology." *New Media & Society* 23, no. 7 (2021): 1899–919.
- Manovich, Lev. *The Language of New Media*. Cambridge and London: The MIT Press, 2001.
- Marantz, Andrew. "Reddit and the struggle to detoxify the Internet.", 2018. <https://www.newyorker.com/magazine/2018/03/19/reddit-and-the-struggle-to-detoxify-the-internet/amp>. Accessed 12 Mar 2018.
- Massanari, Adrienne L. *Participatory Culture, Community, and Play: Learning from Reddit*. New York: Peter Lang, 2015.
- Pedregosa, Fabian, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, and Bertrand Thirion., et al. "Scikit-learn: machine learning in python." *JMLR* 12 (2011): 2825–30.
- Pennebaker, James. *Writing to Heal: A guided journal for recovering from trauma & emotional upheaval*. Oakland, CA: New Harbinger, 2004.
- Ratcliffe, Matthew. *Experiences of depression: A study in phenomenology*. Oxford, UK: Oxford University Press, 2014.
- Ridge, Damien, and Sue Ziebland. "Understanding depression through a 'coming out' framework." *Sociology of Health & Illness* 34, no. 5 (2012): 730–45.
- Romele, Alberto, Marta Severo, Paolo Furia, Romele, A, P. Furia. "Digital hermeneutics: from interpreting with machines to interpretational machines." *AI & SOCIETY* 35, no. 1 (2020): 73–86.
- Sánchez Querubín, Natalia. (2020), "Illness online: popular, tagged, and ranked bodies", PhD diss, University of Amsterdam.
- Seko, Yukari, and Stephen P. Lewis. "The self—harmed, visualized, and blogged: remaking of self-injury narratives on Tumblr." *New Media and Society* 1 (2018): 180–98.
- Slatman, Jenny, and Inge van de Ven. "Gutted: Depression as Embodied Socio-Political Experience in Michel Houellebecq's *Sérotinine*." *Medical humanities* (2020).
- Sontag, Susan. *Illness as Metaphor*. New York, NY: Farrar, Straus and Giroux, 1978.
- . *Regarding the Pain of Others*. New York, NY: Penguin, 2004.
- Sosnoski, James. "Hyper-Readings and Their Reading Engines." In *Passions, Pedagogies, and Twenty-First Century Technologies*, edited by GaileE Hawisher, and CynthiaL Selfe, 161–77. Logan, UT: Utah State University Press, 1999.
- Strawson, Galen. "Against Narrativity." *Ratio* 17, no. 4 (2004): 428–52.
- Strong, Tom. "Conceptualizing interactional patterns: Theoretical threads to facilitate recognizing and responding to IPs." In *Patterns of interpersonal interactions: Inviting relational understandings for therapeutic change*, edited by K Tomm, D Wulff, StGeorge S., and T Strong, 36–56. New York, NY: Routledge, 2014.
- Van de Ven, Inge. "Karl Ove Knausgård and the Serial Self." *Between – Rivista internazionale dell'Associazione italiana di teoria e storia comparata della letteratura* 6." 2 (2016): 31–49.
- . *Big Books in Times of Big Data*. Leiden UP: Leiden, 2019.
- Van der Does, A.J.W, and F.G. Zitman. "Stemmingsstoornissen." In *Handboek Psychopathologie Dl. 1: Basisbegrippen*, edited by W Vandereycken, C. A. L Hoogduin, and P. M. G Emmelkamp, 195–229. Houten: Bohn Stafleu van Loghum, 2008.
- Van Nuenen, Tom. "Text research on online platforms: Heuristic steps and pitfalls." In *Doing Research In and On the Digital. Research Methods Across Fields of Inquiry*, edited by Cristina Costa, and Jenna Condie, 83–101. London: Routledge, 2018.
- Van Nuenen, Tom, and Inge van de Ven. "Digital Hermeneutics and Media Literacy: Teaching the Red Pill across Horizons." *Tilburg Papers in Culture Studies* 201, paper 241.", 2020. <https://www.tilburguniversity.edu/research/institutes-and-research-groups/babylon/tpcs>.
- Wasson, Sara. "Before narrative: episodic reading and representations of chronic pain." *Medical Humanities* 44, no. 2 (2018): 106–12.
- Wohlmann, Anita, and Madaline Harrison. "To be continued: serial Narration, chronic disease, and disability." *Literature and Medicine* 37, no. 1 (2019): 67–95.
- Wellman, Barry. "Physical place and Cyberplace: the rise of personalized networking." *International Journal of Urban and Regional Research* 25, no. 2 (2001): 227–52.
- Wohlmann, Anita. "The Illness Is You: Figurative Language in David Foster Wallace's Short Story 'The Planet Trillaphon.'" In *Embodied Narration: Illness, Death, and Dying in Modern Culture*, edited by Heike Hartung, 203–25. Bielefeld: Transcript, 2018.
- Woods, A.. "The limits of narrative: provocations for the medical humanities." *Medical Humanities* 37, no. 2 (2011): 73–8.
- Zwicky, Jan. *Wisdom and Metaphor*. Kentville: Gaspereau Press, 2008.

Correction: *Digital hermeneutics: scaled readings of online depression discourses*

van de Ven I, van Nuenen T. Digital hermeneutics: scaled readings of online depression discourses. *Med Humanit* 2022;48:335–346. doi:10.1136/medhum-2020-012104

The funding statement has been updated in the online HTML and PDF.



OPEN ACCESS

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

Med Humanit 2023;49:147. doi:10.1136/medhum-2020-012104corr1

