

THE SOCIAL AND PSYCHOLOGICAL WORK OF METAPHOR:
A CORPUS LINGUISTIC INVESTIGATION

by

JANE DILKES

A thesis submitted to
The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

Department of English Language and Applied Linguistics
School of English, Drama, and Creative Studies
The University of Birmingham
September 2021

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

This thesis investigates the triangular relationship between metaphor use, community, and state of mind, to ask the question: what social and psychological work does metaphor do, in the computer-mediated discourse setting of an online forum. The thesis goes beyond the finding and grouping of metaphors for analysis to consider the pattern of metaphor use over time in terms of (i) surrounding language style; (ii) density of use; and (iii) use by different participant groups. In achieving its aim the thesis provides insights into (i) the effect of metaphor use in terms of state of mind; (ii) the role of metaphor in the characterisation of a community; and (iii) methods for considering linguistic metaphor in naturally occurring discourse in terms of its psychological effect, which also creates insights into metaphor theory.

The primary novel contribution of the thesis is to combine an analysis of metaphor use with an analysis of the language style that surrounds it, using established research relating language style to state of mind to consider the social and psychological work that metaphor does. The primary prediction of the investigation is that where metaphor is used to characterise a concept, the surrounding language will be of a style that has been found to be associated with better mental health. This is related to and supported by the second novel contribution of the thesis, which is to consider the role of metaphor in the formation and evolution of a community over time, by considering change in density of metaphor and other key variables in the data as a whole, and for comparative participant groups. The third novel contribution of the thesis is that, alongside more established corpus linguistic techniques, new techniques from the fast-evolving areas of data science and natural language processing are explored and evaluated in terms of (i) finding metaphors in the corpora; (ii) analysing language style; and (iii) diachronic analysis.

It is shown that use of the identified dominant metaphor themes in each community co-occurs with specific language styles associated with mental health, and that this work of metaphor evolves over time as a consensus which becomes normative within the group for a period, such that it shapes community members as well as being shaped by them, while the flexibility of metaphor still leaves that work open to further evolution. The adaptation and prominence of particular metaphor themes over time to do particular work in each forum also underpins the characterisation of it as a particular community.

Acknowledgements

My thanks first and foremost go to my supervisor Jeannette Littlemore, from whom I have learnt so much, and whose expertise and enthusiasm I have found so inspiring. Thank you also to Kate Reid for interesting and helpful discussion about the experience of illness during the first months of this PhD, and to Andy Soundy for super-quick feedback on chapters.

My thanks also to Jim Dilkes for critical technical feedback during the early stages of this PhD, and for being an incredible source of data science expertise throughout; to Silas Dilkes for your enthusiasm, and for helpful discussion around my early diachronic analysis; to Oscar Dilkes for helpful and encouraging feedback on finished chapters, and for checking my calculations; and to Andy for your support, and for always being certain that this PhD was a good thing to do.

Contents

1	<i>Introduction</i>	1
1.1	Introduction to the thesis	1
1.1.1	Aim of the thesis	1
1.1.2	The social and psychological work of metaphor: a first example	3
1.2	Outline of the chapter	7
1.3	Metaphor theory	7
1.3.1	Conceptual metaphor theory	8
1.3.2	Blending theory	15
1.3.3	Deliberate metaphor theory	17
1.3.4	Dynamic approaches to metaphor	20
1.3.5	Metaphor approach in the current investigation	22
1.3.6	Metaphor and health discourse	26
1.3.7	Section summary: metaphor theory	27
1.4	Language style and mental health	28
1.4.1	Function words as a measure of the social and psychological	28
1.4.2	Research linking language style to mental health	29
1.4.3	Social media studies linking language style to state of mind	32
1.4.4	Linguistic Inquiry and Word Count (LIWC)	36
1.4.5	Summary of psychological correlates of linguistic categories	39
1.5	Discursive psychology	39
1.5.1	About discursive psychology	40
1.5.2	Interpretative repertoires	41
1.5.3	Ideological dilemmas	43
1.5.4	Subject positions	47
1.5.5	Emotion discourse	48
1.5.6	Section summary: discursive psychology	51
1.6	Community	52
1.6.1	Metaphor and community	54
1.6.2	Section summary: community	55
1.7	Chapter discussion	56
1.8	Outline of the thesis	60
2	<i>Prospecting for metaphors</i>	62
2.1	Introduction	62
2.2	Cancer and metaphor	62
2.3	The cancer.net data	65
2.3.1	Ethics	65
2.3.2	Collecting data from cancer.net	68
2.3.3	Processing the data	69
2.3.4	Participants and posts: statistics and distribution over time	70
2.3.5	cancer.net as a community	72
2.3.6	Section summary: the cancer.net data	73
2.4	Keywords and key terms	73
2.4.1	Keywords for cancer.net in comparison with a general web corpus	74
2.4.2	Key terms for cancer.net in comparison with a general web corpus	75
2.4.3	Section summary: keywords and key terms	76
2.5	Selecting data based on the number of posts made	77
2.5.1	Section summary: selecting data based on number of posts made	80

2.6	Unsupervised learning introduction.....	80
2.6.1	Automatic metaphor identification.....	80
2.6.2	Natural language processing.....	83
2.6.3	Lexical unit of focus	84
2.6.4	Vectorization	86
2.6.5	Term frequency.....	88
2.6.6	Using k-means clustering to consider the effectiveness of term frequency control for finding metaphors.....	89
2.6.7	Section summary: unsupervised learning introduction	92
2.7	Using dimensionality reduction to find interpretative repertoires.....	93
2.7.1	Principal Component Analysis	94
2.7.2	Factor Analysis	95
2.7.3	Latent Dirichlet Allocation.....	98
2.7.4	Section summary: using dimensionality reduction to find interpretative repertoires.....	99
2.8	Topic modelling with Latent Dirichlet Allocation.....	99
2.8.1	Determining the optimal number of topics.....	101
2.8.2	Visualisation of the LDA output	102
2.8.3	LDA topic modelling output.....	103
2.8.4	Section summary: topic modelling with Latent Dirichlet Allocation.....	110
2.9	Using word vectors to explore metaphoric language	111
2.9.1	Vector plot.....	111
2.9.2	Using vector norms to find dissimilar concepts in a text	114
2.10	Chapter discussion	117
3	<i>Evaluating potential metaphor</i>	119
3.1	Introduction	119
3.2	Metaphor identification procedures.....	121
3.2.1	MIP	121
3.2.2	MIPVU	122
3.2.3	DMIP	123
3.2.4	Discourse Dynamics Framework.....	124
3.2.5	Metaphor identification method used in the current investigation	125
3.3	Investigation into metaphoricality of terms	127
3.3.1	Dictionary definition of terms	128
3.3.2	journey and related terms	129
3.3.3	step and related terms	131
3.3.4	rollercoaster and related terms.....	133
3.3.5	fight and related terms	134
3.3.6	battle and related terms.....	135
3.3.7	Section summary: investigation into metaphoricality of terms	138
3.4	Metaphoric phrases and metaphoremes	140
3.4.1	Graph-based ranking introduction	141
3.4.2	Short example of graph-based ranking	143
3.4.3	Graph-based ranking of the cancer.net.20 data	145
3.4.4	All cancer.net.20 data	146
3.4.5	journey phrases	147
3.4.6	rollercoaster phrases	150
3.4.7	fight phrases.....	152
3.4.8	battle phrases	153
3.4.9	Section summary: metaphoric phrases and metaphoremes	156
3.5	Signalled metaphor	158
3.5.1	Using pattern matching to identify metaphor	160
3.5.2	Hearst patterns	161
3.5.3	spaCy matcher	162

3.5.4	Similes	163
3.5.5	Copular similes	166
3.5.6	Clausal similes: as if / as though	169
3.5.7	so to speak/as it were	173
3.5.8	Superordinate terms: a sort of/a kind of	175
3.5.9	Explicit markers.....	177
3.5.10	Section summary: signalled metaphor	179
3.6	Chapter discussion	181
4	<i>Metaphor and language style.....</i>	185
4.1	Introduction	185
4.2	Supervised learning.....	188
4.2.1	Overview of the process	189
4.2.2	Supervised methods compared	189
4.2.3	Evaluating the models	190
4.2.4	Two different labelling methods.....	191
4.2.5	Considering the number of posts in each condition	192
4.2.6	Section summary: about supervised learning	192
4.3	Text-based method (TEXT)	193
4.3.1	Keras deep learning Python library	193
4.3.2	Text-based model	194
4.3.3	Comparison of different methods	195
4.3.4	Predictions made by the text-based model: all metaphor themes combined.....	205
4.3.5	Section summary: supervised learning text-based method.....	208
4.4	Count-based method (COUNT)	209
4.4.1	Counting lexical items in each post.....	209
4.4.2	Word count: variable counts converted to proportions	215
4.4.3	Selection of supervised machine learning models.....	216
4.4.4	Stratified k-fold.....	217
4.4.5	Count-based method, analysis by labelling type and metaphor theme.....	217
4.4.6	Comparison of results when only pronoun counts are used	219
4.4.7	Comparison with other automatic metaphor identification systems	220
4.4.8	Section summary: supervised learning	223
4.5	Comparing language style between metaphoric conditions	225
4.5.1	journey metaphor theme	227
4.5.2	rollercoaster metaphor theme	232
4.5.3	fight metaphor theme.....	236
4.5.4	battle	240
4.5.5	Section summary: comparing language style between metaphoric conditions	243
4.6	Chapter discussion	245
5	<i>Diachronic analysis.....</i>	248
5.1	Introduction	248
5.1.1	Organisation of the chapter.....	249
5.2	Density of metaphor use over time.....	249
5.2.1	Mean density of each metaphor theme per post, by month	249
5.2.2	Proportion of posts using each metaphor theme by month	252
5.2.3	Section summary: density of metaphor use over time.....	256
5.3	Correlation of metaphor themes with key variables over time.....	257
5.3.1	journey metaphor theme	257
5.3.2	rollercoaster metaphor theme	260
5.3.3	fight metaphor theme.....	263
5.3.4	battle metaphor	266
5.3.5	Section summary: correlation of metaphor themes with key variables over time.....	269

5.4	Use of key variables for comparative participant groups over time	269
5.4.1	Change in density of key pronouns for different participant groups over time.....	270
5.4.2	Change in density of metaphor themes for different participant groups over time	273
5.4.3	Section summary: use of key variables for comparative participant groups over time.....	274
5.5	Diachronic analysis discussion	275
5.6	Cancer.net discussion.....	276
6	<i>suicide.net</i>	283
6.1	Introduction	283
6.1.1	Suicide.net background	285
6.1.2	Suicide and the internet	287
6.1.3	Suicide and language style	289
6.2	The suicide.net data	290
6.2.1	Collecting data from suicide.net.....	290
6.2.2	Processing the data	291
6.2.3	suicide.net participants and posts: statistics and distribution over time	291
6.3	Keywords and key terms.....	293
6.3.1	Keywords: comparison of suicide.net with enTenTen18	293
6.3.2	Key terms: comparison of suicide.net with enTenTen18	295
6.3.3	Comparison of different suicide.net groups with enTenTen18	297
6.3.4	Comparison of keywords between suicide.net participant groups	298
6.3.5	Comparison of key terms between suicide.net participant groups	300
6.3.6	Section summary: Keywords and key terms	301
6.4	Interpretative repertoires	302
6.4.1	Using LDA topic modelling to find interpretative repertoires	302
6.4.2	Using word vectors to explore metaphoric language	307
6.4.3	Section summary: interpretative repertoires.....	309
6.5	Evaluating potential metaphor	310
6.5.1	journey and related nouns.....	311
6.5.2	ctb and related nouns	312
6.5.3	fight and related nouns	313
6.5.4	war and related terms.....	314
6.5.5	Comparative densities of metaphor for different participant groups.....	314
6.6	Metaphoremes and subject positions	316
6.6.1	Graph-based ranking of the suicide.net data	318
6.6.2	journey phrases	318
6.6.3	ctb phrases	321
6.6.4	fight phrases.....	322
6.6.5	battle phrases	324
6.6.6	struggle phrases	326
6.6.7	Section summary: metaphoremes and subject positions	328
6.7	Metaphor and general language style	330
6.7.1	Text-based method	331
6.7.2	Count-based method	339
6.7.3	Section summary: metaphor and general language style.....	340
6.8	Metaphor and specific language style	341
6.8.1	journey	342
6.8.2	ctb	343
6.8.3	fight.....	345
6.8.4	battle	347
6.8.5	Section summary: metaphor and specific language style.....	350
6.9	Diachronic analysis	351
6.9.1	Density of use of metaphor themes over time	351

6.9.2	Correlation of metaphor themes with key variables over time.....	353
6.9.3	Change in density of key pronouns for different participant groups over time.....	358
6.9.4	Change in density of metaphor themes for participant groups over time.....	360
6.9.5	Section summary: diachronic analysis	361
6.10	Signalled metaphor	362
6.10.1	Suicide and metaphor.....	362
6.10.2	Density of signalled metaphor type in different metaphoric conditions.....	364
6.10.3	Analysis of signalled metaphor.....	365
6.10.4	Section summary: signalled metaphor	371
6.11	Chapter discussion	372
6.12	Next chapter	378
7	<i>Discussion and Conclusion</i>	379
7.1	Discussion.....	379
7.2	Conclusion.....	388
7.2.1	Contributions to new knowledge.....	388
7.2.2	Limitations and future work	394

REFERENCES

399

1 INTRODUCTION

1.1 Introduction to the thesis

1.1.1 Aim of the thesis

The aim of this thesis is to investigate the triangular relationship between metaphor use, community, and state of mind, to ask the question: what social and psychological work does metaphor do, in the computer-mediated discourse setting of an online forum. The thesis goes beyond the finding and grouping of metaphors for analysis, to consider the pattern of metaphor use over time in terms of (i) surrounding language style; (ii) density of use; and (iii) use by different participant groups. In achieving its aim, the thesis provides insights into (i) the potential effect of metaphor use in terms of state of mind; (ii) the role of metaphor in the characterisation of a community; and (iii) methods for studying linguistic metaphor in naturally occurring discourse in terms of its psychological effect, which may also create insights into the contested area of metaphor theory.

1.1.1.1 Predictions and novel contributions of the thesis

Metaphor is a powerful social and psychological mechanism, in which a target concept is described in terms of source concepts, such that existing understandings of the source concepts may be made more available to shape understandings of the target concept, a process which entails that other aspects of the target concept are suppressed (Lakoff and Johnson, 2003; Cameron *et al.*, 2009; Müller, 2010).

The current investigation compares two large corpora of naturally occurring, computer-mediated discourse, each consisting of all posts from an active and long-term internet forum established to address a particular life-threatening experience. Forum one,

henceforth for the sake of brevity and anonymity referred to as cancer.net (Hunt & Brookes, 2020), supports discussion around the experience of living with cancer, while forum two, henceforth referred to as suicide.net, is a 'pro-choice' suicide forum, where 'pro-choice' means that suicide is accepted as a valid choice. These two fora were selected because of the potential threat to life inherent in their topic of focus, and for the contrast between discourse relating to a physical health condition (cancer.net) and mental health concerns (suicide.net). Cancer was selected as the physical health condition of focus because of the large body of research considering metaphor in conjunction with the experience of cancer (Gibbs & Franks, 2002; Semino et al., 2017; Skott C, 2002). Suicide.net was selected as the mental health focus because of current debates in the media (Boyd, Halle and Sharpe, 2020; BBC News, 2020; BBC News, 2021) and UK parliament (*Closure of suicide fora - Early Day Motions - UK Parliament*, 2020) about its legitimacy, and calls for it to be banned. Due to the existential challenges inherent in the health conditions that are the focus of each forum it is anticipated that consequently there will be (i) more contrast between posts of language use relating to state of mind (the psychological); and (ii) more evidence of community in language use, including the adaptation of metaphor that is pervasive in the wider culture to characterise, and do particular work within, the community (the social). In addition, internet fora are structured such that it is possible to analyse data for individual participants; interactions between participants; groups of participants; and particular topic threads; and to do all of this diachronically, which is an important aspect of the current investigation.

Building on similar large-scale linguistic investigations of internet fora that have focused on particular health conditions (Semino *et al.*, 2017; Hunt and Brookes, 2020), the primary novel contribution of the current investigation is to combine an analysis of metaphor use with an analysis of the language style that surrounds it, using established

research relating language style to state of mind to consider the social and psychological work that metaphor does. The primary prediction of the current investigation is that where metaphor is used to characterise a concept, the surrounding language will be of a style that has been found to be associated with better mental health. This is an intuitive insight based on research data linking language style to state of mind, which I have linked to conceptual metaphor theory, in which linguistic metaphor is similarly understood as being underpinned by cognitive action. Since use of metaphor to express and negotiate more abstract concepts suggests a higher level of processing and understanding of those concepts, I have surmised that this will be supportive of better mental health. This is related to and supported by the second novel contribution of the thesis, which is to consider the role of metaphor in the formation and evolution of a community over time, by considering change in density of metaphor and other key variables in the data as a whole, and for comparative participant groups. The third novel contribution of the thesis is that, alongside more established corpus linguistic techniques, new techniques from the fast-evolving areas of data science and natural language processing are explored and evaluated in terms of (i) finding metaphors in the corpora; (ii) analysing language style; and (iii) diachronic analysis.

1.1.2 The social and psychological work of metaphor: a first example

This section sets out a first example of the primary prediction of this thesis, that where metaphor is used to characterise a concept it will be surrounded by language of a style that has been found to be associated with better mental health. The following two posts from cancer.net were selected from a search for the term *boxing*, which has been found to be predominantly metaphoric in that context. Post A contains no prominent metaphor, while post B uses a range of metaphoric terms from different conceptual domains, including *journey* and *fight*, with extended metaphor within and across those domains.

For the sake of privacy, post A, which is more personal, is not quoted in full, but the subsequent analysis was applied to each full post. In post B metaphoric phrases are underlined where the source domain is instantiated as a noun, which as discussed later is the focus of the current study.

Figure 1-1 post A (no metaphor) 206 words

on boxing day i found a lump under my right arm, i got a gp¹ appointment at the earliest ... the longer i wait before anything is done, the worse it will get.

Figure 1-2 post B (multiple metaphors) 254 words

so sorry you're going through this heart breaking time at the moment... it's truly a cruel time ... but as someone on their cancer journey... my view on life since getting that and having a total right mastectomy... i live every day ... take all those days as a bonus ... i try to find something every day that makes me smile ... if you live in the day, and build as many memories as you can, then no day is wasted no one is promised tomorrow... anyone healthy or with cancer can be gone ... so holding on to the now .. there maybe things he still wants to do .. or see .. yeah emotions with cancer is like a rollercoaster... up and positive one minute .. crashing down and scared the next ... we all get that ... it's not about being brave ... i think brave is ... being scared witless but you do it any way ... it's about admitting you're all scared ... sharing tears ... listening to each other ... walking side by side ... together ... otherwise everyone tries to be brave ... and that just leads to feeling broken and lonely ... separately.. no do this together ... better or worse ... sickness and health ... you can do this... and if you get knocked down ... give yourself permission to feel whatever emotions you feel ... you're not alone ... lots of us here have been where you are now... then get back up ... and boxing gloves back on.. and help him kick cancers butt along the road ...

¹ In the UK gp, an acronym for general practitioner, is a typical way to refer to a community doctor

There is a notable difference in stance between the two posts, with post B (metaphoric) directed towards the other, offering support, advice, and the solidarity of shared experience, while post A in comparison, which contains no metaphor, is focused on the current lived experience of the participant. That difference in stance comes partly from the use in post B of multiple metaphors, with multiple different source domains, to communicate a lived personal understanding of a difficult and complex experience, as well as to frame that experience in a wider context with a temporal element: to position it beyond the current lived experience of the individual. But in addition to the contrast created by the absence and presence of metaphor, the language style of each post is also very different. For example, post B uses a wider range of personal pronouns (12), to address a wider range of perspectives, a characteristic of language style that has been found to be associated with better mental health. Post A in comparison uses a smaller range of personal pronouns (5), of which three are first-person singular pronouns (marked with asterisks in the table below). The first-person singular pronoun *I*, which as is discussed in more detail later in this chapter has been shown to be a strong indicator of state of mind, such that higher use is associated with worse mental health, has a high density of 5.29% in post A, and a low density of 1.05% in post B, with 3.64% the typical density of *I* in general discourse (Sexton and Pennebaker, 2009).

Table 1-1 comparison of pronoun use between post A (non-metaphoric) and post B (metaphoric)

	pronoun	post A count	post A density	post B count	post B density
1.	he			1	0.35
2.	him			1	0.35
3.	*I	12	5.29	3	1.05
4.	it	5	2.2	4	1.39
5.	*me	3	1.32	1	0.35
6.	*my	9	3.96	1	0.35
7.	she	4	1.76		
8.	us			1	0.35
9.	we			1	0.35
10.	you			9	3.14
11.	your			1	0.35
12.	yourself			1	0.35

The difference in stance and related pronoun use between non-metaphoric post A and metaphoric post B may be related to each participant's level of establishment within the community. In post B the use of metaphor to frame a shared experience communicates a greater authority on the subject, and a greater authority within the community, as well as simultaneously expressing the ethos and normative stance of the community. This also effectively extends an invitation and welcome to the addressee to join that community. The effect of such metaphor is investigated more widely, in more detail, and from different analytic approaches, over the course of the thesis, including consideration of the status and work of different metaphor themes in the data as a whole; for comparative groups of more established and less established participants; and in terms of how they are adopted and adapted over time to characterise the community. This may be related to the findings of Cameron (2007) that metaphor, which may be essential to the expression of unfamiliar ideas and experiences including values, emotions, and attitudes (Cameron, 2007, p. 200), through its intrinsically flexible nature allows ideas to be "challenged, negotiated, and accepted" over time (Cameron, 2007, p. 219).

1.2 Outline of the chapter

The remainder of this chapter addresses the three main areas of analysis of the thesis - metaphor, language style as it relates to state of mind, and discursive psychology - setting up approaches and methods that are used in subsequent chapters. Section 1.2 presents a comparative overview of metaphor theory and concludes by describing the initial approach adopted in the current investigation. Section 1.3 then considers the association of language style with state of mind, summarising key findings that are used in the investigations of subsequent chapters to consider the social and psychological work that metaphor does. Section 1.4 introduces discursive psychology as the overarching analytic approach taken in this thesis, section 1.5 considers method for identifying virtual community, and finally, in section 1.6, the current chapter is discussed as a whole and the content of subsequent chapters is outlined.

1.3 Metaphor theory

Metaphor is a figure of speech in which a target concept is described in terms of source concepts, such that existing understandings of the source concepts may be made more available to shape understandings of the target concept, a process which entails that other aspects of the target concept are suppressed (Lakoff and Johnson, 2003; Cameron *et al.*, 2009; Müller, 2010).

Metaphor is a powerful social and psychological device that fundamentally influences the way we live whether we are consciously aware of it or not. It has many varied functions, including filling lexical gaps; expressing emotional attitudes; decoration, disguise, and hyperbole; cultivating intimacy; re-conceptualisation; and ideology (Goatly, 2011).

Metaphor is also an active research area, with evolving and competing conceptualisations of what is understood as a very powerful social and psychological mechanism. Relevant

metaphor theory will now be discussed for consideration against future insights, and to support the theoretical approach to metaphor that is taken during the current investigation. This starts with conceptual metaphor theory (CMT) and goes on to describe other theoretical frameworks which build on CMT, some of which complement CMT while others are somewhat conflictual.

In every rhetorical definition of metaphor is implied not just a philosophical position, but a conceptual network within which philosophy as such is constituted (Derrida, 1974)

1.3.1 Conceptual metaphor theory

Lakoff and Johnson compiled their influential book introducing conceptual metaphor theory (CMT) *Metaphors We Live By* (Lakoff and Johnson, 2003), first published in 1980, to address claims held widely in the Western philosophy tradition that concepts are conscious, literal, and disembodied. Although the primary insight of CMT that metaphor is part of thought and not just of language appeared at the time of its publication to be radical, Gibbs (1994) shows that this idea has been active for centuries. Conceptual metaphor theory however provided the first definition of (i) what conceptual metaphor is; and (ii) an empirical method for its analysis (Gibbs, 2009). The range of interdisciplinary research relating to CMT suggests that it is the current dominant perspective on metaphor (Gibbs, 2009).

CMT states that metaphor operates at the level of thinking, making the ontological claim that a metaphor links two conceptual areas of meaning, each referred to as a domain, such that the entities that form each domain are connected in the brain. A more abstract target domain may be understood in terms of a source domain, a concept more grounded in perception and experience - the physical and social interactions within an environment -

some of which may be universal, while others may vary between cultures. For example, in the conceptual metaphor LIFE IS A JOURNEY the more abstract target concept LIFE is mapped onto, i.e. takes on the structure of, the commonly experienced source concept JOURNEY. In the terms of CMT since there is no direct way of perceiving more abstract concepts they must necessarily be understood in terms of metaphor, and this directionality from the more grounded to the more abstract is inherent in the theory. And further, linguistic metaphors are possible precisely because of the existence of conceptual metaphors in the brain, such that linguistic metaphors provide insights into the way the more abstract target domain is structured in the minds of their users.

Strong empirical support for CMT comes from a diverse range of methods and across a wide range of research areas (Lakoff and Johnson, 2003; Gibbs, 2009), including:

systematic patterns of conventional expressions across a number of domains and languages (both spoken and signed), lexical generalisations, generalisations across novel cases, historical change, child language acquisition, metaphorical discourse, psycholinguistic findings, and neural computational models of metaphor (Gibbs, 2009, p. 16)

Lakoff and Johnson (2003) argue that the form of language itself reflects conceptual metaphorical meaning. The conceptual metaphor CLOSENESS IS STRENGTH OF EFFECT influences our understanding of literal closeness within forms in a sentence as reflecting the strength of the closeness of the relationship holding between those forms. For example:

1. They don't think it'll leave until tomorrow
2. They think it won't leave until tomorrow

Statement one is weaker, it is argued, because the negative *don't* in that statement is positioned further from *leave*, the item it negates, than is the negative *won't* in statement two.

Another example of the influence of conceptual metaphor on the form of language is the use of more form to represent more content. Based on the conduit metaphor

LINGUISTIC EXPRESSIONS ARE CONTAINERS, increasing the content of such a container is understood as increasing the strength of that linguistic expression. For example, the extended phrase *It is bi-i-i-i-i-g* represents *It* as being bigger than does the non-extended phrase *It is big* (Lakoff and Johnson, 2003, pp. 128–132).

1.3.1.1 The structure of CMT

In the first edition of *Metaphors We Live By* metaphors were categorised as orientational, ontological, or structural. But the 2003 edition states that all metaphors are structural, in that they map structures to structures; all metaphors are ontological, in that they create target domain entities; and many metaphors are orientational, in that they map orientational image-schemas.

CMT describes metaphors as working at various levels of specificity, from AFFECTION IS WARMTH, which works at a primary level based on the association of happiness with the early experience of bodily warmth a child feels when they are close to someone who loves them, to more complex metaphors such as THEORIES ARE BUILDINGS, which have no such perceptually direct explanation, but rather are "grounded in systematic correlations within our experience" (Lakoff and Johnson, 2003, p. 61). A hierarchy is also described within the systematicity of metaphor. For example, the conceptual metaphor TIME IS MONEY entails the conceptually higher-level (more abstract) metaphor TIME IS A VALUABLE COMMODITY. The structure of CMT is summarised as follows:

1. Metaphors are fundamentally conceptual in nature; metaphorical language is secondary.
2. Conceptual metaphors are grounded in everyday experience.
3. Metaphorical thought is unavoidable, ubiquitous, and mostly unconscious.
4. Abstract concepts have a literal core but are extended by metaphors, often by many mutually inconsistent metaphors.
5. Abstract concepts are not complete without metaphors.
6. Our conceptual systems are not consistent overall, since the metaphors used to reason about concepts may be inconsistent.
7. We live our lives on the basis of inferences we derive via metaphor. (Lakoff and Johnson, 2003)

1.3.1.2 The partial nature of metaphorical structuring

The systematicity that CMT describes as allowing us to understand an aspect of a target concept in terms of an aspect of a source target necessarily hides other aspects of the target concept: the relationship between the source and target domains is always one-way, and partial, such that there is always more in the source domain than is carried over to the target domain. This leads to the understanding of metaphor as a powerful tool for framing the world in particular ways that are potentially ideological or rhetorical. Cultural change typically arises from the introduction of new metaphorical concepts and the loss of old ones, it is argued, for example the Westernisation of cultures is described as partly a matter of introducing the TIME IS MONEY conceptual metaphor (Lakoff and Johnson, 2003).

1.3.1.3 Metaphors we live by

when "mouth" applied only metaphorically to bottles, the application made the hearer notice a likeness between animal and bottle openings. (Consider Homer's reference to wounds as mouths.) Once one has the present use of the word, with literal application to bottles, there is nothing left to notice. There is no similarity to seek because it consists simply in being referred to by the same word (Davidson, 1978, p. 37)

Metaphors we live by are the pervasive everyday, automatic, metaphors, which may be understood as being lexicalised or from some perspectives 'dead', with no active metaphoric link being made, in use, between the source and target concepts. For example the conduit metaphor LINGUISTIC EXPRESSIONS ARE CONTAINERS, which frames language as a conduit for information, e.g. I gave you that idea. It is argued that English speakers use the conduit metaphor, which is said to account for around 70% of English metalanguage, automatically, and without recognition that they do so (Reddy, 1979). In the terms of CMT it is these pervasive, invisible metaphors that are alive in the most fundamental way because they literally, and even without us necessarily being aware of it, shape the way that we live. For example, Reddy (1979) argues that there is a frame conflict inherent within the pervasive conduit metaphor that can bias the thought of English speakers, such that it contributes to the social and cultural problems of English-speaking communities.

There is much empirical evidence to show that even apparently instinctive metaphor evokes a connection to the source domain (Katz, 1996; Boroditsky, 2000; Zhong and Liljenquist, 2006; Gibbs and Matlock, 2008). For example, Thibodeau and Durgin (2008) found that reading conventional metaphors, which may typically be considered to be less active, facilitates subsequent understanding of more creative metaphors. In addition,

reading conventional metaphors based on the same conceptual metaphor as the subsequent creative metaphor better facilitates understanding of that creative metaphor than does reading conventional metaphors based on a different conceptual metaphor.

The fact that they are conventionally fixed within the lexicon of English makes them no less alive (Lakoff & Johnson, 2003, p. 55)

1.3.1.4 Deliberate metaphor

In contrast to metaphors we live by, more deliberate metaphor, for example creative or idiosyncratic metaphor, is more aligned with the concept of metaphor that was the focus of earlier philosophical and literary studies - the concept of metaphor that CMT replaced as the dominant research paradigm. In CMT, creative metaphor is defined as occurring in the following three circumstances:

1. extension of the used parts of a metaphor
2. instances of the unused parts of a metaphor
3. instances of a novel metaphor

Although Reddy (1979) argues that changing our currently instinctive use of the conduit metaphor would effect a fundamental change in our interaction with the world, self-conscious potentially deliberate metaphor may be more implicated in re-framing, and ideology. For example, from the pragmatist philosopher Richard Rorty:

the metaphor is an essential tool in the process of re-weaving our beliefs and desires. Without them there would be no such thing as scientific revolutions or a cultural change, but only a change in the truth values of statements, which are formulated in a non-changing vocabulary (Rorty, 1989)

Deliberate metaphor from perspectives other than that described by CMT is discussed in more detail in a subsequent section.

1.3.1.5 Status of CMT

Although CMT is highly influential, and has offered insights into the organisation of conceptual structures, phenomenological experience, and the communicative and aesthetic work of language (Gibbs, 2015, p. 5), it is still a controversial theory.

Disciplinary concerns have led to debate over "whether CMT offers a satisfactory theory of meaning (for philosophers), insights into creativity and poetic practice (for literary scholars), an online account of people's immediate comprehension of verbal metaphor (for psycholinguists), or cultural models (for anthropologists)" (Gibbs, 2017, p. 7).

However, even the more controversial, epistemological, claims of CMT are also still "in line with a fairly broad range of research on metaphor in cognitive anthropology, sociology, and linguistics" (Müller, 2008, pp. 56–57), and have been supported by psycholinguistic experiments.

It has been argued that CMT makes assumptions about what is natural, or normal, such that it has an inherent political stance. For example, Leezenberg (2009) argues that experience is not biologically given, but social from the start, such that to address the fact that the body has a history of power linguistic practices must be given precedence over cognitive considerations. Similarly, although there are reasons to see ANGER IS HEAT for example as the product of the universal physiological correlation between anger and elevated skin temperature, culture is recognised as having a role in the scaffolding and shaping of even such primary metaphors (Grady, 1997): it is recognised that the humoral theory of emotions probably also played a role in the development, conventionalisation, and elaboration of this metaphoric pattern in Western languages (Grady, 2010).

Casasanto (2013, 2017) similarly suggests that primary metaphors may be shaped by culture, for example the form of writing systems, as much as by embodiment. The related question of variation in bodily experience is addressed by Littlemore (2019) in a wide ranging, empirically supported discussion of sources of variation in embodied metaphor, which questions the CMT assumption that universal bodily experiences underpin conceptual metaphor.

Gibbs (2009) notes that some criticisms of CMT come from influential commentators whose research has little to do with metaphor, and may typically be focused on the original edition of *Metaphors We Live By* (Lakoff and Johnson, 2003), without taking into account the vast range of research and discussion that has helped to evolve the theory of conceptual metaphor. For example, primary metaphor theory (Grady, 2010), in which primary metaphors are described as simple patterns which map fundamental perceptual concepts onto equally fundamental concepts that are not directly perceivable, such as HAPPY IS UP. Although primary metaphor theory does not completely explain the reason some words and not others are metaphorically mapped from source to target domains, it "does provide a crucial limit on why some metaphorical constructions are likely to occur, and others not" (Gibbs, 2017, p. 30).

Other metaphor theory will now be discussed that builds on CMT to address perceived gaps in its explanatory power.

1.3.2 Blending theory

Blending theory is concerned with conceptual integration in which multiple conceptual structures are combined (Fauconnier and Turner, 2002). It is a broad theory that attempts to explain meaning making more generally but has been applied extensively to the study of metaphor, in particular it has been argued that blending theory is better able to support

an understanding of creative metaphor, while CMT better represents more conventional metaphor (Kövecses, 2015). Where CMT is concerned primarily with conventional patterns of association, i.e. patterns which we can think of as objects stored in long-term memory, blending is in principle a dynamic process. It takes place when a thought or its expression uses more than one input space, where input space is defined as a fluid mental space structured by a conceptual domain. Although CMT and blending theory have been understood as competing theories (Littlemore, 2019), it has also been argued that in addressing different aspects of the conceptualisation of metaphor they are mostly complementary (Grady, Oakley and Coulson, 1999). The different intuitions about metaphor that CMT and blending theory represent may reflect different cognitive styles. For example, a test of holistic in comparison to analytic thinking styles found that holistic thinkers, who organise information into 'chunks', were significantly more likely to blend conceptions of source and target domains than were analytic thinkers, who process information as separate parts (Boers and Littlemore, 2000).

Metaphoric blends arise, it is argued, when one input space contributing to a metaphor is structured by the source domain of an existing conceptual metaphor, and another input space contributing to the metaphor is structured by the target domain of that metaphor. Because blending analyses take the form of linked networks of representations, they are not directional in the same sense as CMT mappings, thus they allow for feedback from a blended space to any of its inputs (Fauconnier and Turner, 2002).

Blending theory does not entail that all metaphors are blends. For example, the metaphor *prices rose* is an example of the conceptual metaphor MORE IS UP, a spatial rise in which no specific entity is rising in space, such that no blended target domain is invoked (Lakoff and Johnson, 2003). Example one below represents such a conventional mapping,

with the policies of a state represented by the course of a ship. Example two in contrast, which introduces the additional concepts of stormy seas and a safe harbour, is a more complex interaction of concepts and links. By supporting multiple spaces and iterations, such that one blend may be the input for another, it is argued, the blending framework helps to capture such a complex interaction (Grady, Oakley and Coulson, 1999).

1. Without the consent of our fellow citizens, we lose our moral authority to steer the ship of state.
2. The [Sri Lankan] ship of state needs to radically alter course; weather the stormy seas ahead and enter safe harbour.

Alongside CMT, blending theory is a widely accepted account of metaphor, which may be essential to the explanation of how complex mixed and extended metaphor may be understood (Gibbs, 2001), and has also been found to be an effective explanation for aspects of metonymy that are otherwise difficult to explain. For example the RESULT FOR ACTION metonymy *digging your own grave*, in which the act of digging your own grave followed by the act of dying are compressed into a single phrase, must be unpacked into a blended space in order to be understood (Littlemore, 2017, p. 60).

1.3.3 Deliberate metaphor theory

As discussed above, CMT does address deliberate metaphor. However CMT is typically understood as focusing on *metaphors we live by*, the pervasive, instinctive and often invisible metaphors that dominate language to shape the way we live, that from some perspectives are considered to be 'dead', such that there is no active metaphoric link between the source and target concepts. It has been argued that the CMT focus on instinctive metaphor has led to the neglect of the powerful communicative aspect of metaphor (Steen, 2008, 2011, 2017; Deignan, 2011); and of explanations based on an

individual's previous language experience, and their beliefs, attitudes, emotions, and values (Cameron and Deignan, 2006). It is important to address these gaps because although metaphor is a fundamental part of everyday language, it is also a vital mechanism for communicating the most intense human experiences that are otherwise impossible to convey (Deignan, 2005).

The three dimensional model of metaphor is a discourse-analytic approach that introduces a communicative aspect to metaphor, alongside language and thought, to address the question whether or not a particular use of metaphor is deliberate (Steen, 2008).

This framework is not cognitive-linguistic or psycholinguistic, but rather discourse-analytical. When metaphor is studied as part of actual language use, or events of discourse, it does not only manifest a linguistic form and a conceptual structure, but also a communicative function (Steen, 2008, p. 221)

Consideration of the communicative aspect of metaphor, it is argued, addresses a paradox of metaphor: that most metaphor in language is not processed metaphorically, i.e. is not processed by a cross-domain mapping involving some form of comparison, as opposed to within-domain categorisation (Bowdle and Gentner, 2005). Consideration of the communicative aspect of metaphor also reinstates the traditional distinction between metaphor as a tool for rhetoric, versus metaphor as a tool for more general concerns of language and thought, while still being grounded in the achievements of CMT (Steen, 2008). From this perspective there are three functions of metaphor, naming, framing, and perspective changing, as follows:

1. The linguistic function of metaphor is to fill lexical (and other formal) gaps in the language system (metaphor in language); we may want to call this naming.

2. The conceptual function of metaphor is to offer conceptual frameworks for concepts that require at least partial indirect understanding (metaphor in thought); we may want to call this framing.
3. The communicative function of metaphor is to produce an alternative perspective on a particular referent or topic in a message (metaphor in communication); we may want to call this perspective changing, or simply changing. (Steen, 2008, p. 231)

Deliberate metaphor theory (DMT) is an evolution of these ideas, stating minimally that intentional use of metaphors as metaphors implies that attention is paid to the source domain as a separate domain of reference to the target domain (Reijnierse *et al.*, 2018).

The career of metaphor framework (Bowdle and Gentner, 2005), an empirical investigation into how metaphoric mappings are established between concepts from different domains, has shown that whether metaphors operate at the level of individual concepts (comparisons) or entire conceptual domains (categories), depends on both their degree of conventionality and their grammatical form, with more conventional metaphors more likely to be instantiated as categories, and less conventional metaphors as comparisons. In terms of grammatical form, similes were found to invite a comparison of the target domain to the source domain (comparison), while metaphors promote the classification of the target domain as a member of a category named by the source domain (categorisation) (Bowdle and Gentner, 2005). An fMRI investigation into the neural career of metaphor supports the approach, finding that the conventionalisation of novel metaphors "entails a honing of the neural networks within a region as opposed to across regions" (Cardillo *et al.*, 2012, p. 12). There is a difference, then, between the structure and instantiation of more and less conventional metaphor, and the response that may be invoked by these different types of metaphor.

There have been concerns that DMT undermines the dramatic and productive shift that CMT made away from the previously dominant understanding and analysis of metaphor as a creative linguistic device, with its focus on creative and artificial (non-naturally situated) examples. But accusations that "DMT takes us back to a Stone Age time where metaphor was ornamental, deviant, and only employed by special people with highly conscious communicative aims" (Gibbs and Chen, 2017, p. 124) are based on differing interpretations of specific terms such as consciousness, and communication, it is argued (Steen, 2017); and are a misleading contrast of deliberateness with conventionality (Müller, 2011); such that rather than returning the focus to the expressive, DMT cuts across a distinction between expressive and explanatory analogy (Deignan, 2011). DMT is compatible with current metaphor paradigms such as CMT, it is argued, while revealing crucial, powerful new aspects, such as the role of intentions and attention (Steen, 2017). This view is supported by Zlatev *et al.* (2021), who set out desiderata for a contemporary metaphor theory, including that it must account for communication as well as cognition.

1.3.4 Dynamic approaches to metaphor

understanding a metaphor is as much a creative endeavour as making a metaphor, and as little guided by rules (Davidson, 1978, p. 31)

In dynamic approaches to metaphor it is argued that rather than focusing on a distinction between deliberate and instinctive metaphor, a concept of metaphor is needed which addresses metaphors as processes (Müller, 2011). Use of the more neutral concept of activation, rather than deliberateness or consciousness, it is argued, allows integration of the cognitive, affective, and interactive processes that are involved in the creation and activation of metaphor (Müller, 2011). In a dynamic approach, the linguistic and

cognitive realms of metaphorical structure are both recognised as important and as having the potential to be active, with conceptual metaphor understood as being accessed through the specific semantics of linguistic metaphor. From this perspective metaphor is not dead or alive, or automatic or deliberate, but sleeping or waking, and which of these is the case is determined by cognitive activation in a particular individual, including those who express the metaphor and those who receive the metaphor, at a particular moment in time.

1.3.4.1 Discourse dynamics framework

The discourse dynamics framework (DDF) is an approach that understands metaphorical language and metaphorical thinking as affecting each other in the dynamic and dialogic processes of talking and thinking, in which talking is understood as a fundamental part of thinking. In the CMT approach discussed above, metaphorical expressions in language are seen as manifestations of conceptual metaphor, which are held to be prior in three ways:

1. in thought as prior to language
2. across speech communities as prior to individuals
3. in more general form as prior to specific instantiations

In the dynamic approach in comparison no priority is given to thought over language, or to the general over the specific. Since from this perspective metaphor can't be separated from its discourse context without becoming something different, so the connections between discourse context and metaphor use must be described and explained, such that in DDF consideration is given to:

1. connections between metaphors and the discourse context where they are used

2. connections across metaphors in a particular discourse event
3. connections between metaphors used in a particular discourse event and metaphors used more broadly across society and over time (Cameron and Maslen, 2010)

In this way the dynamic approach addresses some empirical concerns that have been raised about CMT, for example:

1. Why are linguistic metaphors apparently subject to grammatical and lexical restrictions?
2. If linguistic metaphors are the expression of a broad conceptual mapping, why are they so unevenly and inconsistently distributed?
3. Why do different languages and cultures make different use of source domains? (Cameron and Deignan, 2006)

As well as these differences, however, CMT and DDF are each based on the same understanding that connected patterns of metaphors, such as systematic metaphor, are important tools in understanding and discourse (Cameron *et al.*, 2009).

1.3.5 Metaphor approach in the current investigation

CMT remains a dominant theory in metaphor research, such that other current influential approaches such as deliberate metaphor theory, and the dynamic approach, discussed above, are all underpinned by CMT, having been developed to address specific gaps and limitations. Blending theory has been discussed as another widely accepted account of metaphor, which may be essential to the explanation of how complex mixed and extended metaphor may be understood. Current theories have been set out individually so that as the current investigation progresses, results may be linked back to theory, potentially providing insights into this ongoing debate. For example, it is anticipated that the

language style surrounding particular metaphor themes will change over time, for individual participants, in the data as a whole, and in the context of different uses at different times. This would support a dynamic view of metaphor, while also, since language style is in the current study taken to be associated with state of mind, linking this to cognition. In addition, the anticipated evolution and adaptation of pervasive metaphor such as LIFE IS A JOURNEY to support the evolving ethos, ideology, and rhetoric, within a community, would support the understanding of that metaphor as having a specifically communicative or deliberate aspect, which is itself variable. The focus in the current study is on such potentially more deliberate metaphor, since it is less likely that more pervasive instinctive metaphor, i.e. *metaphors we live by*, such as the conduit metaphor (Reddy, 1979), may be corralled to have particular work hitched to it within a community.

Gibbs (2009, p. 17) describes seven broad questions about CMT that remain to be addressed, including the issue of particular practical relevance to the current corpus linguistic study: "How does one decide what counts as evidence for conceptual metaphor?" Steen (1999) sets out five steps from linguistic to conceptual metaphor, as follows:

1. metaphor focus identification
2. metaphorical idea identification
3. non-literal comparison identification
4. non-literal analogy identification
5. non-literal mapping identification

Deignan (2017) argues that this framework does not address the issue of how one particular conceptual metaphor rather than another can be identified on the basis of language data. For example, the linguistic metaphor 'I think you all deserve a medal' may be a realisation of a journey mapping, or fighting in a war, or competing in an art or dog show (Cameron, 2003). There is also the even more fundamental issue of whether a linguistic metaphor is a realisation of any conceptual metaphor at all. However, linguistic metaphor has historically been used as evidence for conceptual metaphor from CMT onwards. And since there is currently no satisfactory method of identifying conceptual metaphor in text the current investigation necessarily starts with a focus on linguistic metaphor. The focus in the current study on the language style surrounding metaphor use as it is associated with state of mind creates a connection from linguistic metaphor to cognition that has not been explored before. It may potentially support a methodology for identifying conceptual metaphor from linguistic metaphor, which remains a pressing issue for metaphor research (Gibbs, 2017).

The primary practical focus of the current investigation, then, is the ontology-neutral concept of metaphor themes, where a metaphor theme is "an abstraction from the metaphorical statements in which it does or might occur. A metaphor-theme is available for repeated use, adaptation and modification by a variety of speakers or thinkers on any number of specific occasions" (Black, 1977, p. 438). The related concept metaphoreme, which may be part of a wider metaphor theme, is typically a smaller semantic and pragmatic unit of analysis that combines "the linguistic, the cognitive, the affective, and the socio-cultural" (Cameron and Deignan, 2006, p. 686) in a particular speech community at a particular time. For example, in a class of school students the metaphorical phrase *lollipop trees* became part of the repertoire "at least for a while" as a way to refer to the unacceptability of trees drawn in a particular simplistic manner

(Cameron and Deignan, 2006, p. 677). In a study of the use of the metaphoreme *the cancer card* in an online community relating to cancer it was noted that although metaphoremes may also be observed at the level of a whole speech community, they are particularly useful when accounting for specific discourse communities (Semino and Demjén, 2017), which is a primary focus of the current study.

Metaphor themes may also be understood in terms of the discourse dynamics framework analytic concept of systematic metaphor, which is defined as successive episodes and instances of vehicle development around closely connected topics, leading to an emerging group of closely connected (systematic) metaphors, for example 'journeys' (Cameron, 2007). Systematic metaphors represent a temporary shared language resource in the dynamic process of thinking and talking, which may constrain how discourse participants think and talk about topics, but which may also evolve further as discourse continues. Empirical study of metaphor dynamics has shown that various changes and adaptations may be made to metaphors as discourse proceeds, with metaphor shifting identified as a key mechanism that leads to the emergence of systematic patterns of metaphor use. Metaphor shifting includes, for example, the same or semantically connected items being re-used with a different topic; the vehicle term being repeated, relexicalized, clarified, or contrasted; and shifting between metaphorical and literal uses which may lead to words and phrases carrying symbolic meaning even when they are used in non-metaphorical ways (Cameron, 2007; Cameron *et al.*, 2009). The later diachronic analysis of the current study considers the dynamic aspect of identified metaphor themes, such as is represented by the analytic concepts of systematic metaphors, and metaphor shifting.

Although the initial approach of the current investigation is necessarily ontologically neutral, such that the focus is on linguistic metaphor instantiated as metaphor themes, and

metaphoremes, the cognitive, communicative, and dynamic aspects discussed above are also addressed in the wider investigation into the language style surrounding metaphor use, and the pattern of metaphor use over time and between participant groups. It is necessary to start from a pragmatic, empirically firm position, even while still recognising the dynamic nature of metaphor, since metaphor analysis can become vacuous if all metaphor is denoted as dynamic (Zlatev, Jacobsson and Paju, 2021). Methods for identifying linguistic metaphors in text are discussed in chapter three, in which potential metaphor from cancer.net is evaluated.

1.3.6 Metaphor and health discourse

In terms of the health focus of the current investigation, metaphor in conjunction with the experiences of living with cancer, and suicidal ideation, is discussed in chapters two and six respectively alongside the investigations into those specific areas. Consideration is now given to the more general use of metaphor in conjunction with health discourse.

In terms of healthcare professionals, metaphor may be used to communicate concepts that are technically or emotionally difficult to understand, in what for patients are typically novel experiences (Czechmeister, 1994; Skott C, 2002; Kirklin D, 2007), with health professionals who use analogies and metaphors rated as having better communication skills than those who do not (Casarett *et al.*, 2010). For example a survey study to critically interrogate typical assumptions of psychotherapists about the therapeutic functions of metaphor suggests that metaphoric language is perceived to be more effective than is literal language in the discussion of clients' concerns (Tay, 2020). Metaphor contributes to the creation of common language such that a consensual understanding can be established in the therapeutic relationship (Reisfield and Wilson, 2004). And an argument is made for a coherent framework of metaphor use to convey

complex new information, and that this is particularly important in health campaigns (El Refaie, 2015).

Metaphorical skill, imagination, and sensitivity are important in creating rapport and in communicating the nature of unshared experience (Reisfield and Wilson, 2004, p. 4027)

In terms of people living with illness, metaphor may be used to communicate emotional, physical, and social, aspects of an experience which is likely to be atypical for them and for those to whom they are close (Bowker, 1996; Gibbs & Franks, 2002). Metaphor may be used to exercise control through the choice of imagery it entails (Bowker, 1996).

Metaphor preference may also be different in different contexts and at different times, and for different purposes, since no single metaphor may be sufficient to communicate and understand the experience of severe illness over time (Williams Camus, 2009).

Metaphor is an important aspect of health discourse, then, and as such it should be an area of active social and psychological work within the two health-related corpora of the current investigation.

1.3.7 Section summary: metaphor theory

In this section current metaphor theory was considered. It was established that the current text-based corpus investigation will necessarily start with an ontologically neutral, linguistic approach. However, analysis of the social and psychological work of metaphor through consideration of its surrounding language style and evolution over time is expected to draw on and connect to all of the theory described above, including blending theory, deliberate metaphor theory, and the dynamic approach. The current study may also support new methods for identifying conceptual metaphor from linguistic metaphor, which is a pressing issue for metaphor research (Gibbs, 2017).

In the next section, the next key method of analysis for the investigation is considered, which is the association of language style with state of mind. At the end of the section, key findings are summarised for use in subsequent chapters in which investigations are made into the social and psychological work that metaphor does.

1.4 Language style and mental health

The analysis of subtle shifts in pronouns, articles, and other almost-invisible words can reveal the psychological effects of life experiences (Seraj et al., 2021, p. 1)

The analysis of language style to provide insights into mental health is an active area of research that is particularly suited to the diachronically organised, naturally occurring personal data on the many social media platforms that are an increasingly integral part of our social and psychological lives. In addition, methods have developed from word counting by hand (Weintraub, 1981) to computerised word counting in a wide range of linguistic categories (Pennebaker *et al.*, 2015) to the application of new techniques from the rapidly evolving research areas of natural language processing and machine learning, an approach which is explored in the current investigation. In this section, the association of language style with mental health is considered, concluding with a summary of key findings that are used in subsequent chapters to support analysis of the social and psychological work of metaphor.

1.4.1 Function words as a measure of the social and psychological

Function words, including pronouns, articles, prepositions, auxiliary verbs, negations, conjunctions, quantifiers, and common adverbs, are a fundamental aspect of language style. Content words, which label an object, or actions, include nouns, verbs, adjectives, and adverbs. From a psychological perspective, function words reflect how people are

communicating, while content words in comparison convey what they are saying: function words are much more closely linked to people's social and psychological worlds (Pennebaker, 2011).

Although English speakers have on average a vocabulary of almost 100,000 words, only about 500 (0.05%) of these are the function words that connect, shape, and organize content words, while at the same time function words make up around 55% of words used (Pennebaker, 2011). Function words do not carry much intrinsic meaning and are typically not noticed, and in natural language processing they are often specifically excluded as 'stop words'.

1.4.2 Research linking language style to mental health

The psychiatrist Walter Weintraub, an early practitioner of word counting to locate relationships between language and health, noticed that elevated use of first-person singular pronouns is linked to elevated levels of depression (Weintraub, 1981). Since then many studies have used word counts to consider how language style is associated with mental health; a few of the studies more relevant to the current investigation are described in more detail here.

1.4.2.1 Word use in the poetry of suicidal and non-suicidal poets

In an analysis of approximately 300 poems from the early, middle, and late, period of nine suicidal poets and nine non-suicidal poets, it was found that the poetry of suicidal poets contains, throughout their careers, more first-person singular pronouns, and fewer first-personal plural pronouns (Wiltsey Stirman and Pennebaker, 2001).

In another study of texts by suicidal and non-suicidal poets using the LIWC (Linguistic Inquiry and Word Count) software, predictors of suicide were found to include use of a

higher percentage of self-oriented words, negative emotions, ambiguity, and an attitude of excluding reality, plus a lower proportion of words related to others, world entitativity, and experiencing the world (Pająk and Trzebiński, 2014).

1.4.2.2 Language use of depressed and depression-vulnerable college students

Analysis of essays written by currently depressed, formerly depressed, and never depressed college students showed that currently depressed students use more negatively valenced words, and more of the first-person singular pronoun I, than do never depressed students. Formerly depressed students do not differ from never depressed participants overall, however, formerly depressed students do increase their use of I across the essays, such that in the final set of essays their use of I is significantly greater than it is for never depressed students (Rude, Gortner and Pennebaker, 2004).

1.4.2.3 Expressive writing studies

In a study of three to five sessions of expressive writing completed over several days or weeks, with a maximum of one session per day, it was found that people who experience health benefits from expressive writing tend to show:

1. a high rate of positive emotion words
2. a moderate rate of negative emotion words
3. an engagement with cognitive processing words

It was suggested that the person whose writing positively impacts on their health is expressing a high rate of positive emotion, but also acknowledging and addressing the negative, while the use of cognitive words demonstrates an effort to understand events.

The following writing patterns in comparison were found to be suggestive of worse mental health:

1. a high rate of negative emotion words was linked to use of alcohol and tobacco
2. a higher than average use of the first-person singular pronoun I of around 6% or more may indicate depression, with average use of I being around 3.64%
(Sexton and Pennebaker, 2009)

In another expressive writing study it was found that increased use of causation words, and decreased use of negative emotion words and first-person singular pronouns, predicts an improvement in mental health as reflected in an increase in self-distancing and a decrease in emotional reactivity, and this was found to lead to fewer physical symptoms (Park, Ayduk and Kross, 2016).

1.4.2.4 Analysing language in suicide notes and legacy tokens

In an analysis of 25 suicide notes, 21 legacy tokens of active shooters (homicide), and a sample of 20,000 general student texts, it was found that negative emotion is used twice as often, and anger four times as often, in the homicide texts than in the suicidal or general student texts, with no difference in use of these items between the suicidal and general student texts. Personal pronouns are nearly three times as prevalent in the suicidal texts as they are in the other texts, and the future tense is nearly four times as prevalent, while there is no difference in use of these linguistic items between the homicidal and general student texts.

Because there is no difference between the two other conditions for linguistic items that are significantly different in each case (suicidal, homicidal), it was surmised that the significantly different variables in each condition "tap into a deeper cognitive construct that is unique to internal or external harm orientation" (Egnoto and Griffin, 2016, p. 145).

1.4.2.5 When small words foretell academic success: the case of college admissions essays

While this study may seem less relevant to the current thesis, the 'analytic thinking' factor it describes is referred to in a subsequent highly relevant study.

In a factor analysis of function word categories on 25,000 college admission essays a single factor was found in which articles and prepositions are positively loaded, and the remaining categories are negatively loaded. This 'analytic thinking' factor was found to be positively correlated with 4-year grade point average (Pennebaker *et al.*, 2014).

1.4.3 Social media studies linking language style to state of mind

Social media allows a diachronic consideration of language style using large amounts of naturally occurring personal data. Research using social media data has found language patterns relating to a range of emotional and psychological states, including the onset of depression (De Choudhury, Counts and Horvitz, 2013; Yates, Cohan and Goharian, 2017; Eichstaedt *et al.*, 2018; Guntuku *et al.*, 2019); suicidal ideation (Choudhury and Kıcıman, 2017; O'Dea *et al.*, 2017; Desmet and Hoste, 2018; Van den Nest, Till and Niederkrötenhaler, 2019; Schoene *et al.*, 2021); and post-traumatic stress disorder (Coppersmith, Harman and Dredze, 2014).

The following recent English-language based social media studies focus on language style to predict state of mind, predominantly without considering specific relevant content words (for example *suicidal*, or *depressed*). A German-language based study of suicide fora is also included because of its high relevance to the analysis of suicide.net in chapter six of the current investigation.

1.4.3.1 A linguistic analysis of suicide-related Twitter posts

A set of Twitter posts expert coded for suicide risk was compared with matched non suicide risk Twitter posts. Strongly concerning suicide related posts were found to have a higher word count, increased use of first-person singular pronouns, and more references to death in comparison to non suicide related posts; they had an increased use of first-person singular pronouns, greater use of anger words, and an increased focus on the present in comparison to non concerning suicide related posts (O'Dea *et al.*, 2017).

1.4.3.2 Mental distress and language use: linguistic analysis of discussion forum posts

In a study of online peer support groups for people experiencing Generalized Anxiety Disorder, Borderline Personality Disorder, Major Depressive Disorder, Obsessive-Compulsive Disorder, or Schizophrenia, people who experience any of these conditions were found to have a higher frequency of first-person singular pronouns and a higher frequency of negative emotion words compared to a control group. Those participating in schizophrenia discussion fora were found to have the highest use of third person plural pronouns, while those in the Borderline Personality Disorder fora were found to have the highest use of third person singular pronouns, which is suggested to be due to an excessive focus on significant others, alongside possible insecure attachment patterns (Lyons, Aksayli and Brewer, 2018).

1.4.3.3 Comparing indicators of suicidality among users in different types of non-professional suicide message boards

Analysis of 1,200 threads from seven German language non-professional suicide fora found fewer emotion words and words related to social circumstances, and a higher rate of death- and aggression-related words, in pro-suicide fora. It was noted as a limitation

that complex language structures can not be analysed with the word counting method used (LIWC) (Van den Nest, Till and Niederkrotenthaler, 2019).

1.4.3.4 Linguistic markers in women's discussions on miscarriage and abortion illustrate psychological responses to their experiences

In a study of thousands of Reddit conversations relating to miscarriage and abortion experiences, classified as unplanned and planned trauma respectively, women experiencing planned trauma were found to use more self-distancing language patterns*, and to engage in more emotion regulation, while women who experienced unexpected trauma were found to use more self-focused and social-based language (Blackburn *et al.*, 2021). Jaeger *et al.* (2014) similarly show that those who use fewer self-focused words after a trauma tend to experience better psychological adjustment.

*Self-distancing can be seen in a higher use of third person pronouns and impersonal pronouns; for example, people with depression tend to use fewer third person pronouns compared to a control group.

1.4.3.5 Hierarchical multiscale recurrent neural networks for detecting suicide notes

Noting that people were increasingly posting their last words online, this investigation focused on blog posts in three categories: Genuine Suicide Note data (GSN), Depression Notes from the Reddit depression data (DN1 and DN2), and Neutral Blog Posts (NEU1 and NEU2). Using a recurrent neural network, a prediction rate F1 score of .96 over the baseline of 0.6 was achieved for a condition in which data is unbalanced, in that there is a vast amount of neutral data, which reflects real-world scenarios (Schoene *et al.*, 2021). There was found to be a greater use of function words in the non-neutral data, with an increase in use of the first-person singular pronoun I, and a greater use of negations. In

addition, in terms of content word categories, verbs were found to be increased in the non-neutral data in comparison to the neutral data.

Table 1-2 prevalence of function and content words in contrasting conditions

word category	GSN	NEU1	NEU2	DL1	DL2
function	56.80	47.87	49.12	58.27	59.35
personal pronoun	15.85	10.06	10.23	14.35	14.32
I	10.64	6.45	6.26	11.63	11.60
negation	2.87	1.47	1.65	3.10	3.24
verb	19.06	16.46	15.92	21.10	21.40
adjective	4.54	4.71	4.25	4.80	4.82
adverb	4.79	5.27	5.64	6.91	7.20

For the GSN (genuine suicide notes), pronouns, positive emotions, and an increased focus on the present, were the most important linguistic categories for accurate classification. In DL2 (Reddit depression data) references to death ('I'm dying inside'), work ('unemployment'), negative emotions, and an increased focus on the past, were assigned the highest attention weights by the neural network. In NEU2 (neutral blog posts) in comparison there were fewer personal pronouns, increased use of adjectives and adverbs, and fewer emotion references (Schoene *et al.*, 2021).

1.4.3.6 Language left behind on social media exposes the emotional and cognitive costs of a romantic breakup

In a diachronic analysis of all posts across multiple subreddits of participants on the Reddit internet forum r/BreakUps thread, the sharpest drop in 'analytic thinking' (Pennebaker *et al.*, 2014) was found to occur at the time of disclosure of a relationship breakup, while language was found to be at its most personal and informal just after the breakup, even when participants were not talking about their relationship (Cohen's $d=0.758$). At the same time as *analytic thinking* dropped, there was found to be an increase in *cognitive words*, *I-words* and *we-words*, with *I-words* having the longest period of increased use. This pattern was apparent across all subreddits and the effect was

apparent, although less strong, even when data from subreddits relating to breakups and relationships were excluded. Statistically significant pattern changes were present from one month before the breakup up to 3.5 months after, peaking at the time of the breakup with Cohen's *d* values of 0.270 (*cognitive words*), 0.580 (*I-words*), and 0.407 (*we-words*).

Cognitive words and *we-words* were found to return to baseline values within a month after breakup disclosure, while *analytic thinking* and *I-words* took 14 and 10 weeks respectively to return to baseline values (Seraj, Blackburn and Pennebaker, 2021).

1.4.4 Linguistic Inquiry and Word Count (LIWC)

The well established LIWC program (Pennebaker *et al.*, 2015) used in some of the studies discussed above counts words in a wide range of semantic and linguistic categories, with results displayed as a proportion of the total words. The number of categories and distribution of words between them has evolved through use: the most recent version of LIWC, LIWC2015, has over 80 linguistic, psychological, and topical categories (Pennebaker *et al.*, 2015). LIWC can not be used to analyse metaphor directly, but in the current study research findings from LIWC regarding the psychological correlates of language style are used to consider the psychological work of identified dominant metaphor themes of each community, by considering the language style that surrounds their use. LIWC is not used as a method in the current study.

More than a hundred published articles use LIWC in a wide range of experimental settings to demonstrate the psychological implications of particular word use, including "attentional focus, emotionality, social relationships, thinking styles, and individual differences" (Tausczik and Pennebaker, 2010, p. 24). The following table summarises LIWC research findings with specified psychological correlates relating to mental health,

including the number of published papers listed as representing each finding (Tausczik and Pennebaker, 2010).

Table 1-3 summary of psychological correlates of linguistic categories

lexical category	example terms	psychological correlates	papers listed
first-person singular	i, me, mine	honesty, depression, low status, personal, emotional, informal	38
first-person plural	we, us, our	detached, high status, socially connected to the group	23
second person	you, your	social, elevated status	9
third person singular	she, her, him	social interests, social support	13
third person plural	they, their, they'd	social interests, out-group awareness	10
affective processes	e.g. happy, cried, abandon	emotionality	19
negations	e.g. no, not, never	inhibition	9

Application of the LIWC method has shown language style to be accurately associated with, and reflective of, a person's state of mind. However, there are recognised limitations of LIWC, such as that it can't analyse complex language structures (Van den Nest, Till and Niederkrotenthaler, 2019), and that it does not take into account the communicative context of the words that it counts, treating texts simply as "bags of words" (Hunt and Brookes, 2020, p. 238). Pennebaker *et al.* (2015) describe LIWC as a transitional text analysis program; they anticipate new techniques that will apply to the massive diverse natural language data of the internet to relate patterns of language style with content words, and consider the structure of text beyond single words. For example, Support Vector Machines focusing on stemmed bigrams, and the absence or presence of features, rather than their rate of use, have outperformed LIWC in a large scale task to classify the personality of bloggers (Iacobelli *et al.*, 2011). The current investigation does not use LIWC but does use the many relevant findings from LIWC to support analysis

using other text analysis methods from the active and evolving fields of natural language processing and data science.

Only a single application of the LIWC method used in conjunction with metaphor analysis was located in the literature. In that research, a depression measure was developed based on typical metaphors for depression, in conjunction with presence of first-person pronouns, and five other LIWC categories intuitively selected as relevant, for diagnostic screening of online texts for depression (Neuman *et al.*, 2013). From the web, metaphors were collected in which the target term (for example 'depression') was embedded. Using a search engine, web pages were sought that contain the expression '(the term) is like *', where * is a wildcard, a technique based on signalled metaphor, which is explored in chapters three and six of the current investigation. It was found that the vast majority of phrases located in this way are metaphors that perform a "deep structure mapping between concepts" (Neuman *et al.*, 2012, p. 21).

The nouns from the metaphoric phrases found were loaded into the Corpus of Contemporary American English (CoCA) to find potential synonyms, while for verbs the lemma was entered to find potential synonyms. CoCA was then used to design paraphrases in which nouns in the phrases selected were substituted with their synonyms. In this way a lexicon was created of 1,723 metaphorical phrases about depression for use to locate metaphors relating to depression in online texts. Although the depression measure created was not as accurate as psychologists in locating depression in the same texts, it was considered to be a useful preliminary step. Comparing the system's prediction to the judgment of human experts an average 78% precision and 76% recall were achieved (Neuman *et al.*, 2013). This supports the investigative method of the current investigation, which combines a word counting approach with metaphor analysis.

1.4.5 Summary of psychological correlates of linguistic categories

The following recurrent insights relating language style to mental health are considered in future analysis in the current investigation of language style in relationship to metaphor use.

Table 1-4 summary of key findings associating language style with mental health

lexical category	occurrence	associated mental health
first-person singular	increased	worse: depression; self focus
first-person plural	increased	better: more socially connected to the group
third person pronouns	increased	better: social support; self distancing
range of pronouns	diverse	better: considering multiple perspectives
positive words	increased	better
negative words	increased	worse
emotion words	increased	emotionality
negations	increased	inhibition; suicidal ideation

These established findings also provide a platform from which to explore the psychological correlates of other linguistic variables, and combinations of linguistic variables. In addition to these findings, in the current study natural language processing techniques are used to consider a wider range of linguistic categories, with 143 linguistic variables considered in subsequent analysis, as well as a machine learning analysis of whole post texts in relation to metaphor use, which addresses some limitations of the LIWC method by considering the wider structure of discourse. It also addresses a gap in research.

1.5 Discursive psychology

In a corpus study of internet fora relating to mental health it was argued that the combination of 'principled' quantitative techniques with theory-sensitive qualitative techniques offers both a greater insight into the data, and also findings that have a clearer implication (Hunt and Brookes, 2020). Similarly, in a forum-based corpus study of

metaphor, cancer, and the end of life, it was noted that "corpus linguistics makes it possible to combine qualitative and quantitative analysis in an explicit and rigorous manner, and is therefore ideally suited to bridge the methodological divide in healthcare research" (Semino *et al.*, 2017, p. 13). In the current investigation, therefore, quantitative analysis based on counts and co-occurrence of linguistic features in various categories, and their change over time, is combined with qualitative analysis, using a discursive psychology perspective, which also supplies the overarching theoretical approach for the study.

1.5.1 About discursive psychology

Discourse analysis is the close study of language in use, where language is understood as the site at which meanings are created and changed, and not as something that is transparent, or reflective, or a psychological process. In discourse analysis, in order to understand what work language is doing, consideration is given to its situated use within the ongoing process of interaction (Yates, Taylor and Wetherell, 2001).

Whereas conversation analysis (Sacks, Schegloff and Jefferson, 1974) is only concerned with what individual participants produce in any interactional sequence, discursive psychology is a form of discourse analysis that recognizes that people draw on a repertoire of terms from their situated cultural history. In discursive psychology in addition the focus is on psychological themes:

There are three major strands in DP. These are: (i) respecification and critique of psychological topics and explanations; (ii) investigations of how everyday psychological categories are used in discourse; (iii) studies of how psychological business (motives and intentions, prejudices, reliability of memory and perception, etc.) is handled and managed in talk and text, without having to be overtly labelled as such (Edwards, 2005, p. 11)

Discursive psychology is a social constructionist perspective, taking language as its topic rather than understanding language as providing clues to cognition. It rejects the distinction between words and action, understanding language itself as a form of practice. The three main tenets of discursive psychology are that discourse is:

1. constructed and constructive
2. situated
3. action-oriented (Wiggins and Potter, 2017)

Discursive psychology, then, is consistent with the linguistic analysis of metaphor which, as described above, is necessarily the initial analytic approach to metaphor of the current investigation. The three key analytic concepts of discursive psychology are interpretative repertoires; ideological dilemmas; and subject positions; which are discussed in more detail below.

1.5.2 Interpretative repertoires

Interpretative repertoires are relatively coherent recurring systems of terms used to characterise and evaluate actions, events, and other phenomena, that are often organised around specific metaphors and figures of speech (Potter and Wetherell, 1987, p. 149; Edley, 2001). Interpretative repertoires may be understood as part of a culture's common sense, or folk psychology. But whereas much effort has been put into dismissing folk psychology as an inaccurate, inconsistent, and even dangerous theory of mind (Stich, 1983; Churchland, 1988), in discursive psychology "the folk psychology thesaurus has its own reality as the actual terms used by people to perform the actions done in and through everyday discourse" (Edwards, 2004, p. 570).

Compared to the Foucauldian concept of discourses (Foucault, 1971), interpretative repertoires do not apply to entire institutions: they are smaller, and more fragmented, offering speakers a wide range of rhetorical opportunities in any situation. However, in any particular situation a particular rhetoric, or rhetorics, may be preferred, or hegemonic (Gramsci, 2007): they assume the status of facts, such that people become the products of discourse as well as producers of it (Edley, 2001, p. 190).

1.5.2.1 Metaphor and interpretative repertoires

Although metaphors are part of the definition of interpretative repertoires and are understood as characterising the repertoires of which they form a part, in practice metaphors are not typically a key focus of analysis in discursive psychology. However, as has been discussed above, metaphors are an important part of health discourse, being used to communicate emotional, physical, and social, aspects of a difficult experience, and to provide a communicative bridge between professionals and patients. So key interpretative repertoires of the communities under consideration here are expected to contain metaphor, and it is those interpretative repertoires that are the focus of the current study.

The concept of systematic metaphors discussed above in relation to the discourse dynamics framework is also relevant to the concept of interpretative repertoires, in that they both represent a temporary shared language resource in the dynamic process of thinking and talking. The systematic metaphors of a community also may constrain how discourse participants think and talk about topics, such that they are themselves constructed by as well as constructing such metaphor, which may evolve further as discourse continues in the community.

1.5.3 Ideological dilemmas

Ideological dilemmas are a rich and flexible site of social interaction, deliberation, and evolution (Edley, 2001); they "permit the possibility not just of social dilemmas but of social thinking itself" (Billig *et al.*, 1988, p. 17). If lived ideologies are the beliefs, values, and practices, of a given society or culture, ideological dilemmas are the contradictory strands that are prevalent within both ideology and common sense or folk psychology. The tensions and indeterminacy inherent in ideological dilemmas may be understood as enabling meaningful thinking about the self and the world (Billig *et al.*, 1988).

In comparison to formal intellectual ideologies, in which ideological dilemmas are not tolerated, and must be repaired if the speaker is not to appear incoherent and lose credibility, for example by using TWOD - a "truth will out" device (Potter and Wetherell, 1987, p. 153), lived ideologies are not at all integrated or coherent; rather they are characterized by inconsistency, fragmentation, and contradiction (Edley, 2001). But it is precisely that incoherent nature of lived ideologies that makes them a site of change and interaction.

1.5.3.1 Metaphor and ideological dilemmas

The same abstract concept can be metaphorically structured in different ways ... choice of metaphor can have far-reaching ideological as well as cognitive consequences (Goatly, 2011, p. 77)

The analytical concept of ideological dilemmas has specific relevance for the investigation of metaphor. As has been discussed above, metaphor does not exactly represent a literal concept, because in that case it would be that literal concept. Metaphor instead is a mechanism that accentuates some parts of a concept while hiding other parts of the same concept, such that ideological dilemmas are inherent within metaphor. The

structural definition of conceptual metaphor theory (Lakoff and Johnson, 2003), discussed above, includes ideological dilemmas as follows:

- Abstract concepts have a literal core but are extended by metaphors, often by many mutually inconsistent metaphors.
- Our conceptual systems are not consistent overall, since the metaphors used to reason about concepts may be inconsistent.

For example, from the master metaphor list (Lakoff, Espenson and Schwartz, 1991) the illness metaphor TREATING ILLNESS IS FIGHTING A WAR sets up the target concept ILLNESS to be understood in terms of the source concept FIGHTING A WAR. This overarching conceptual metaphor is described as encompassing other potential perspectives on illness, which may not be coherent with each other, for example: 'disease is an enemy'; 'the body is a battleground'; 'the immune system is a defence'; 'being defeated is dying'. In addition, another metaphor from the master metaphor list, LIFE IS A JOURNEY, special case one in the event structure mapping LONG-TERM PURPOSEFUL ACTIVITY IS A JOURNEY, which may also be used in conjunction with the experience of illness, has a very different framing to the fight metaphor. For example, a large corpus study of metaphor, cancer, and the end of life (Semino *et al.*, 2017) found that in comparison to violence metaphors, which are best for expressing defiance, intense activity or concentration, and determination to succeed against the odds, journey metaphors can frame the experience of illness in ways that emphasise continuity, companionship, agency, purpose and potentially positive aspects and emotions. There is also less potential for feelings of anxiety and guilt associated with use of the journey metaphor in relation to illness, particularly when the illness is incurable. However, a journey metaphor may also reflect and reinforce a sense of powerlessness, loneliness, and

the inability to express negative emotions such as frustration, dejection, and lack of acceptance.

Research into metaphor use in conjunction with illness suggests that multiple, possibly contradictory, metaphors may be used at different times, in different contexts, as people work to make sense of and communicate their disrupted experience and existence (Gibbs and Franks, 2002), and this also supports the use of the concept of ideological dilemmas in metaphor research. In addition, multiple contrasting metaphors are used in single posts in the data of the current investigation, which was also found to be the case in the similar investigation by Semino *et al.* (2017). Mixing metaphor in this way is consistent with the blending and dynamic theoretical approaches, while the deliberate metaphor approach, with its focus on the communicative aspects of metaphor, identifies metaphor mixing as specifically deliberate, in that the metaphor draws attention to itself as a stylistic or rhetorical method through the combination of multiple source-domain referents (Steen, 2016). Conceptual metaphor theory in comparison specifically refers to mixed metaphors belonging to two different conceptual domains as impermissible. For example the phrase 'The content of the argument proceeds as follows' (Lakoff and Johnson, 2003, p. 95) combines JOURNEY and CONTAINER conceptual metaphors in a way that can not satisfy both of those concepts at once. However, it has also been argued that CMT naturally entails the use of mixed metaphors in natural discourse (Kövecses, 2016).

The mixing of metaphors, then, may represent the ideological dilemmas experienced by individuals living through and creating a response to existentially threatening experiences. Different metaphor may also be mixed to emphasise the complexity, or intensity, of an experience. For example, a corpus study of interviews with people living with chronic pain found that:

Metaphor mixing occurs where the speaker's purpose is to emphasise the intensity of the embodied experience by representing the pain as out of control. The greater the semantic divergence of metaphor source domains, the more intense the embodied experience of pain, and the greater the agency of the pain rather than the speaker. (Charteris-Black, 2016, p. 155)

In terms of the fora investigated in the current study, the mixing of metaphor may also in some cases represent a practice of more established participants, as the local experts in that specific life experience, to instinctively offer up a range of approaches to other participants to whom they are providing emotional and social support. It is perhaps an automatic, natural implementation of an endeavour represented more formally in the metaphor menu for people living with cancer (Semino, 2019), which arose out of the research project cited above (Semino *et al.*, 2017) to provide a range of metaphor options which may be more or less suitable for particular individuals at particular times.

Considering ideological dilemmas as a specific site of work and evolution of meaning is also consistent with the talking and thinking analytic approach to metaphor represented in the concepts of metaphor shifting, and systematic metaphor, discussed above in conjunction with the discourse dynamics framework (Cameron, 2007; Cameron *et al.*, 2009). It can also be related to the concept of metaphor clustering, in which the density of metaphor use increases over a relatively short span in a text or discourse, which has been found to indicate sites of intensive work relating to the purpose of a discourse (Cameron and Stelma, 2007).

General and pervasive metaphor may be established in a community as a way of conceptualising and communicating new or difficult concepts, for example from scientific discourse. However, while technical use of metaphor tends to be highly conventionalised and specific, when such metaphor is adopted in more general discourse

it may be used and adapted for less technical and specific purposes, which may be deliberate, in ways that represent a misleading and inconsistent conceptualisation (Deignan, Semino and Paul, 2019). The evolution and adaption of metaphor that is in general use in the wider community, such that its use within the specific communities under investigation is not consistent with that wider use, also represents an ideological dilemma and is understood in the current study as particularly revealing of the ethos of the community. This is consistent with the idea of vague language as "a linguistic unit (word, phrase or utterance) that has an unspecified meaning boundary, so that its interpretation is elastic in the sense that it can be stretched or shrunk according to the strategic needs of communication" (Zhang, 2011, p. 573), where metaphor is understood as vague language that may be used to target and realise communication goals. The importance of such change in meaning is recognised by Prandi (2012), who argues that now that the fundamental role of consistent, shared metaphorical concepts in cognition and language is well considered and established, attention must be paid to "the most typical and revealing instances of living metaphor", (Prandi, 2012, p. 148) which are not "poetic rewordings of shared and consistent metaphorical concepts" (Prandi, 2012, p. 148) but contextual re-interpretations of the conflictual meanings inherent in complex metaphor.

Conflictual metaphors document the creative potential of linguistic expressions and shed light on the structure of metaphorical transfer and projection (Prandi, 2012, p. 148)

1.5.4 Subject positions

In discursive psychology, subject positions are identities made relevant by particular ways of talking, including a person's own talk. They connect the wider interpretative repertoires to the social construction of particular selves. This aspect of discursive

psychology is very relevant to the current study which has a focus on language style, with a particular focus on personal pronouns, as being associated with mental health.

Discursive psychology was the theoretical approach used by Horne and Wiggins (2009) to consider the creation of identities of the self and the other on an internet suicide forum, where the identity of being authentically suicidal within a suicide community may be difficult to negotiate. For example, frequent use of *I* was identified as important in presenting "a particular individualised identity, working up emotions and mental states as a consequence of reportedly unbearable external situations" (Horne and Wiggins, 2009, p. 175), while membership entitlement was found to be demonstrated by use of the pronouns *you* and *we* addressed to other participants, which also orientates the speaker as an expert within the forum (Horne and Wiggins, 2009).

1.5.4.1 Metaphor and subject positions

The Marxist philosopher Althusser argued that the way that people experience and feel about themselves and the world around them is a by-product of particular ideological or discursive regimes, the framings of which are typically instantiated in and supported by metaphor. Ideology creates subjects by drawing people into particular positions or identities, such that in subjectification people are both produced by and subjected to ideology (Althusser, 2001). For example, in a large-scale study of metaphor, cancer, and the end of life, both journey and fight metaphors were found to support the construction and negotiation of new identities for people whose identities are likely to have been disrupted, and even destroyed, by their illness (Semino *et al.*, 2017).

1.5.5 Emotion discourse

In addition to the discursive psychology concepts described above, emotion terms are another analytic focus of the current investigation. Although it is part of folk psychology

to consider emotion terms as fixed definitions from cognitive science, a folk understanding which is itself a part of the work that emotion terms do in discourse, in the current investigation emotion discourse is treated not as scientifically fixed, but as productive and flexible:

The rich variety and situated uses of emotion words and metaphors suggest a set of rhetorical affordance in which different parts or potentials of meaning, even contrasting ones for the same word, may be worked up and deployed (Edwards, 1999, p. 271)

As with all discourse, then, so with the use of emotion terms: mental states are understood in the current investigation as the categories and concerns of talk, rather than talk being understood as caused by mental states. And it is specifically the contrast between this and the folk psychological understanding of emotion that allows emotion talk to perform powerful psychological and social work.

It is specifically because people's emotion displays (thus categorized) can be treated either as involuntary reactions, or as under agentive control or rational accountability, as internal states or public displays, reactions or dispositions, that emotion discourse can perform flexible, accountability-oriented, indexically sensitive, rhetorical work (Edwards, 1999, p. 288).

Analysis in subsequent chapters quantifies use of emotion words in various categories and considers how that use changes over time, and in conjunction with other linguistic and semantic categories, including metaphors. In addition, as with the relationship between subject positions and personal pronouns discussed above, emotion discourse has also been shown to have particular meanings in terms of state of mind. For example, increased use of negatively valenced terms has been found to be associated with anxiety

and depression, while use of fewer emotion references has been associated with neutral blog posts in comparison to those of the depressed or suicidal (Schoene *et al.*, 2021).

1.5.5.1 Metaphor and emotion

It is well established that emotion language is dominated by metaphorical and metonymic expressions (Kövecses, 2008). For example, Fainsilber and Ortony (1987) found that descriptions of emotional states contain more metaphorical language than do descriptions of behaviours, indicating that metaphorical language does the important work of expressing what is difficult to express using literal language alone. In addition, intense emotions were found to be associated with greater metaphor use than were mild emotions for descriptions of feeling states, but not for descriptions of actions associated with emotions. Fetterman *et. al* (2016) found that participants more likely to use metaphors have a better understanding of their own emotions; and Semino (2011) found embodied metaphor to be more prevalent in the expression of deeply felt experience, such as pain. The metaphoric representation of emotion, then, is inherent in folk psychology, and therefore an essential aspect of the analysis of emotion displays.

Kövecses (2000) argues that emotion talk is largely metaphorical because metaphor is required to express the variety of diverse and intangible emotional experiences. In terms of the emotionally challenging subjects of cancer and suicide that are the focus of the current investigation, metaphors may be fundamentally important in providing a mechanism for discussion, for example: "by transforming one thing into another they can enable patients to acknowledge their situation without confronting and reflecting on its unbearableness ...Metaphors allow cancer patients to approach the truth obliquely ... and then to distance themselves in a way perhaps less threatening to them, and to the nurse" (Lanceley & Clark, 2013, p. 194). Kövecses (2000) notes, however, that there are no

emotion-specific metaphors, although some emotion source domains are specific both to particular emotion concepts and to the emotion domain. This specificity occurs by means of regular metonymic processes in that we understand an emotion concept via its cause or effect, it is suggested. But although there are no emotion-specific metaphors, emotion metaphors largely fall under the more generic CAUSES ARE FORCES metaphor, based on the force dynamics system which provides much of our folk theory of the human mind (Talmy, 1988).

Another aspect of metaphor in relation to emotion that is important to the current study is the idea that emotions may play a role in the adaptation and evolution of metaphor to specific purposes over time in the online communities under investigation, which were founded to support common deep emotional needs.

the affective—the beliefs, attitudes, values, and emotions of participants—plays a central, but often neglected, role in the emergence of particular forms of metaphor. (Cameron and Deignan, 2006, p. 674)

1.5.6 Section summary: discursive psychology

In the current study, the discourse analytic approach of discursive psychology is used to support qualitative analysis and provides the overarching theoretical approach. The key analytic techniques of discursive psychology are interpretative repertoires, ideological dilemmas, and subject positions, and it has been shown that metaphor is strongly connected to each of these. In addition, emotion discourse is considered as a site of social and psychological rhetoric, and as supporting community adaptation of metaphor to support their specific emotional and social needs.

1.6 Community

In analyzing linguistic phenomena within a socially defined universe, however, the study is of language usage as it reflects more general behavior norms. This universe is the speech community: any human aggregate characterized by regular and frequent interaction by means of a shared body of verbal signs and set off from similar aggregates by significant differences in language usage. (Gumperz, 1968, p. 66)

Alongside consideration of interpretative repertoires, ideological dilemmas, subject positions, and emotion discourse, in each forum analysed in the current investigation the issue must also be addressed of whether the forum is a community, since this has relevance for all of these analytic methods. For example, Morgan (2014, p. 2) states that "it is within speech communities that identity, ideology and agency are actualized in society". Speech in a speech community comes to form a system through the ongoing action of a shared set of social norms. The linguistic characteristics of a speech community are the result of a consensus that evolves in response to changing attitudes. And "Individuals are accepted as members of the group to the extent that their usage conforms to the practices of the day" (Gumperz, 1968, p. 70).

In terms of virtual communities, Herring (2004) warns that there is often the impulse to categorise online phenomena in broader terms than are warranted by their actual instantiation online, "for example, all groups of people interacting online are 'communities'" (Herring, 2004, p. 339). In addition, the increasing prevalence of interaction online means that the distinction between offline and online communication and community is increasingly blurred (Li, 2020). However, there are typical characteristics of internet communication that are substantially different to offline communication, for example increased use of acronyms, and the shortening of text

(Morgan, 2014). Another difference is that for computer-mediated communication (CMC), "the effect of overlap and incomplete, redundantly-initiated exchanges can be likened to a chaotic cocktail party in which every conversation is taking place, equally loudly, in the presence of every guest" (Herring, 1999). It is suggested that the popularity of CMC despite this chaos derives, for example, from the opportunities it provides for increased interactivity and language play. And although turn-taking and topic maintenance may be compromised in CMC, users employ adaptive strategies, for example addressing a specific respondent by name, or referring to a specific message via screen capture (Li, 2020).

Suler (2005) argues that there is a disinhibition effect of online communities that is a result of dissociative anonymity, invisibility, asynchronicity, solipsistic introjection, dissociative imagination, attenuated status, and authority, but with anonymity, invisibility, and asynchronicity having by far the biggest effect. This may be of particular relevance for the fora considered in the current investigation, for example Grohol (1998) found that individuals experiencing mental health issues talk more openly about their experience in such anonymous environments because they remove the fear of negative evaluation, and Gilat and Shahar (2007) similarly found that participants experience an intense sense of affiliation and emotional support through asynchronous communication with the online group. The asynchronous nature of discourse may make it more meaningful, in that more time may be spent in constructing and considering posts (Harvey, 2014); it also gives the option to 'disappear' from an interaction if it becomes too difficult to deal with (Suler, 2005). In addition, the internet provides enhanced opportunities for discussion and connection about quite specific topics, and between people who may have limited opportunities for connection in the non-virtual world, which may lead to increased self-disclosure and increased support for fellow sufferers

who have through their discourse exhibited their membership of the relevant group (Horne and Wiggins, 2009).

The following table summarises discourse behaviours hypothesized to indicate virtual community (Herring, 2004). This set of characteristics is used in the current study to determine whether each of the fora under consideration is a virtual community.

Table 1-5 discourse characteristics of virtual communities

community attribute	behaviours
Structure	Jargon, references to group, in-group/out-group language
Meaning	Exchange of knowledge, negotiation of meaning (speech acts)
Interaction	Reciprocity, extended (in-depth) threads, core participants
Social behaviour	Solidarity, conflict management, norms of appropriateness
Participation	Frequent, regular, self-sustaining activity over time

1.6.1 Metaphor and community

metaphors help to perform some of the central positive functions associated with peer-to-peer online communities, including for people with cancer (Semino et al., 2017, p. 154)

While metaphor is not a part of the standard definition of virtual community described in the previous section, in metaphor research, metaphor is recognised as central to the language conventions that emerge as being characteristic or defining of groups (Cooper, 1989). For example, Veale, Shutova and Klebanov (2016, p. 110) note that "identification of its (a community) common metaphors can teach one about the perceptions of various stake-holders as well as about the dominant theories of a domain, which are often metaphorically structured".

Within-group metaphors have a key purpose in sustaining intimacy among group members and marking the identity of groups through language that is obscure or inaccessible to outsiders. At the same time, as well as being fundamental to

understanding group discourse, being allowed access to within-group language becomes an important marker of access to the group itself. Metaphor can also develop a framing around the specific topic with which the group is associated (Marková, 2007).

In the current investigation, metaphor is understood as playing a fundamental role in characterising community, and in characterising and supporting its interpretative repertoires and ideological dilemmas. Metaphor is understood as a malleable tool that may be recontextualised and honed within a community such that its use within that community may diverge from the typical wider cultural use. It may frame those concepts that have particular relevance within that community in a way that best suits their evolving needs, which simultaneously reinforces the sense that it is a particular community with its own language and meaning. In the current investigation, in addition to the widely used virtual community criteria described in the previous section, then, the following instantiations of metaphor also are considered to be important characteristics of a virtual community:

1. the presence of metaphor in community interpretative repertoires
2. the recontextualisation of prevalent metaphor from the wider culture to perform particular work in the community
3. the suppression within the community of metaphors that may be typical in the folk psychological discourse of the wider culture
4. the creation and persistence within the community of new metaphor that is not used in the wider culture

1.6.2 Section summary: community

Virtual communities have fundamentally different communicative methods and attributes than do physical communities, such that they may encourage more thoughtful,

meaningful, and accessible discussion. While metaphor is not part of the established criteria for identifying virtual community, it is recognised within its own related research area as being an important characteristic of community discourse and the rhetorical work it performs. In the current study, then, in conjunction with established criteria, metaphor use is another consideration in the identification of virtual community.

1.7 Chapter discussion

In this chapter the aims and contributions of the current investigation were discussed, before consideration of the key theoretical areas of the thesis, which include metaphor theory; the relationship between language style and mental health; and discursive psychology as the overarching theoretical approach of the thesis. In addition, theory relating to virtual community was considered, with a set of criteria established against which to consider the fora under investigation, to determine whether they are or are not a community.

Although these areas are well established, in the current literature they are not typically considered together. By combining these approaches the current study aims to create new theoretical insights. For example, while currently metaphor analysis is not combined with a consideration of surrounding language style as an insight into mental health, this has direct relevance for the unresolved issue of how linguistic metaphor may be understood as being underpinned by cognitive action. Consideration of metaphor in conjunction with word counting approaches related to language style is likely to be productive in this regard and may potentially lead to further more focused research in this area. This specific consideration of metaphor in terms of the language style surrounding it also supports the primary research question of this thesis: where metaphor is used to characterise a concept is the surrounding language of a style that has been found to be

associated with better mental health, with the prediction stated above that this will be found to be the case.

As discussed in section 1.5, a qualitative approach combined with the quantitative focus of the current study may support increased insights into the data and findings that have a clearer implication. The social constructionist discursive psychology perspective adopted here as the overarching theoretical approach is coherent with the ontologically neutral, linguistically focused approach to metaphor with which as discussed in section 1.3.5 the current study must necessarily start. In addition, it has been shown above that the three key analytic tools of discursive psychology, interpretative repertoires, ideological dilemmas, and subject positions, are all strongly connected to metaphor.

While in discursive psychology metaphor is not a primary focus, it is recognised within discursive psychology that metaphor may typically characterise interpretative repertoires. In the area of metaphor research in comparison, metaphor is more specifically recognised as being an important characteristic of community discourse and the rhetorical work it performs. In the current study, therefore, it is metaphoric terms from identified interpretative repertoires that are the specific focus. Those key metaphoric terms from interpretative repertoires are analysed for their level of actual metaphoric use, and this is compared for groups of participants who are more established and less established in the community. This approach supports the second research question of the thesis: what is the role of metaphor in the formation and evolution of a community over time. This research question is also related to the discursive psychology concepts of (i) ideological dilemmas as a site of particular social work and change, where as discussed in section 1.5.3.1 ideological dilemmas may be inherent in metaphor, and (ii) subject positions within the community, where particular ideological or discursive regimes of a community may

typically be instantiated in and supported by metaphor (section 1.5.4.1). The second research question of this thesis, which as discussed above is more specifically supported by the analytic techniques from discursive psychology, is also related to and supportive of the primary research question.

The analytical concept of ideological dilemmas is also related in the current study to the mixing of metaphors from different domains in a single post, or to address the same interpretative repertoire, and the mixing of metaphor is also related to deliberateness (sections 1.5.3.1, 3.2.5). Steen (2016) argues that the perception that metaphors are being mixed depends on the interpretation or experience of those metaphors as deliberate: "all other cases of conceptual clashes between adjacent metaphors do not get recognized as mixed metaphor because they are not used deliberately as metaphors" (Steen, 2016, p. 114). From a dynamic perspective (section 1.3.4) this may be related to activation levels of metaphor, where "the mixing of metaphors changes the semantic salience structure, creating different version and degrees of activated metaphoric meaning" (Müller, 2016, p. 32). In comparison to the discourse dynamics approach (section 1.3.4.1), which has a focus on potential metaphoricity of linguistic metaphors for participants in a conversation, Müller (2016) relates the mixing of metaphor to display by participants of the presence of metaphoric meaning in their communicative intent. In terms of the career of metaphor theory (Bowdle & Gentner, 2005), referred to in section 1.3.3 in support of deliberate metaphor theory, the mixing of metaphor perhaps reinstates cross-domain mapping (comparison) when in metaphor in comparison to simile there is typically within-domain mapping (categorisation), which may be more conventional, and therefore potentially less deliberate. The mixing of metaphor may also be related to blending theory, discussed in section 1.3.2, in which multiple conceptual structures are combined, which may be essential to the explanation of how complex mixed and extended metaphor

may be understood, and to CMT, which states that mixed metaphor is impermissible (section 1.3.1). While findings of the current study about a particular effect of metaphor mixing might be considered to undermine the CMT position, as argued in section 1.3.2 in addressing different aspects of the conceptualisation of metaphor, blending theory and CMT are largely complementary.

The word counting approach to consider language style also underpins the diachronic analysis of chapter 5 (cancer.net) and section 6.9 (suicide.net), which supports both the primary and secondary research questions. In considering the change in density of use of key variables related to language style over time, including their use in conjunction with the metaphor themes, and the change in density of metaphor use over time and for comparative participant groups, the current study provides insights into the dynamic nature of metaphor (section 1.3.4). As will be discussed further in sections 2.6.3.4, 3.2.5, and 7.2.1, the discourse dynamics framework (section 1.3.4.1) is not compatible with the predominantly word counting approach of the current study in which, as will be discussed in chapter two, the focus is on metaphor instantiated as nouns in individual phrases, sentences, and posts. However, considering ideological dilemmas as a specific site of work and evolution of meaning (section 1.5.3.1) is consistent with the talking and thinking analytic approach to metaphor represented in the discourse dynamic framework concepts of metaphor shifting, and systematic metaphor (Cameron et al., 2009; Cameron, 2007).

While distinctive use of metaphor is not a formal part of established criteria for determining virtual community (section 1.6), as discussed above metaphor is recognised within its own research area as being an important characteristic of community discourse and the rhetorical work it performs. The current study, then, will use current standards for

identifying virtual community to consider whether each forum investigated is a community, and will also consider whether patterns of metaphor use, and associated surrounding language style, may more specifically support the identification of virtual community. These established and metaphor-related considerations of whether each forum is a community will also support the primary and secondary research questions of the thesis.

Using corpus linguistic techniques in conjunction with exploratory natural language processing and data science techniques, then, the current study combines all of the analytic approaches described in the current chapter to consider what the social and psychological work is that metaphor does in an online community, with the primary prediction, as stated at the start of this chapter, that in general where metaphor is used to characterise concepts in any forum post, the language surrounding it will be of a style that has been associated with better mental health, and where such metaphor is not used in any forum post, the language in that post will be of a style that has been associated with worse mental health. The overarching focus is on the changing work of metaphor as it becomes adapted to specific community purposes, with reference to the idea that re-purposed conceptual metaphors from the wider culture are active sites of social and psychological work in a particular community.

1.8 Outline of the thesis

Chapters two to five of the thesis investigate the cancer.net corpus, using corpus linguistic and data science techniques to find (chapter two) and evaluate (chapter three) metaphors, before considering the language style that surrounds them (chapter four), and their use over time (chapter five). The cancer.net data are discussed as a whole at the end of chapter five, with conclusions drawn about their combined meaning.

In chapter six the techniques used to investigate cancer.net over chapters two to five are applied to the suicide.net corpus, culminating in a discussion of the data. The techniques applied to both cancer.net and suicide.net (chapter six) include: unsupervised learning and word vector analysis to find metaphor (chapter two); concordance analysis to confirm the metaphoricity of identified potentially metaphoric terms (chapter three); consideration of the work of metaphoric noun phrases and metaphoremes (chapter three); consideration of signalled metaphor and its relation to the metaphor themes (chapter three); supervised machine learning to validate the prior method to identify metaphor themes, based on the accuracy of supervised models in predicting the presence in individual posts of the identified themes (chapter four); more detailed consideration of language style in relation to metaphor themes by comparing densities of 143 linguistic variables in posts that do and do not contain each metaphor theme (chapter four); and diachronic analysis to consider the change in metaphor use and key linguistic variables related to state of mind, over time, and for comparative participant groups based on level of participation in the forum (chapter five). Explanation of these techniques and why they were selected is included in chapters two to five as they are applied to cancer.net. The subsequent application of these techniques to suicide.net is undertaken to consider the more general applicability of techniques used to investigate cancer.net, including comparison of results between a corpus relating to predominantly physical health concerns (cancer.net) and a corpus relating to predominantly mental health concerns (suicide.net). The techniques are not discussed in detail when they are applied to suicide.net in chapter six, which is consequently a condensed iteration of the techniques that covers a single chapter.

Finally, in chapter seven results are discussed as a whole, with conclusions drawn about the thesis under investigation, the methods used in the investigation, and what the results may say about metaphor theory.

2 PROSPECTING FOR METAPHORS

2.1 Introduction

In this chapter corpus linguistic and data science techniques are used to locate prevalent metaphoric terms and potential metaphor themes in the cancer.net corpus. The chapter starts with an introduction to the corpus, including how it was created, and its dimensions, moving on to an exploration of its keywords and key terms. A subsequent intra-corpus comparison between more established and less established participants supports the use of data from more established participants only in this phase of the investigation, in which the most typical metaphors of this community and their associated interpretative repertoires are sought. Following this, the active research area of automatic metaphor identification is introduced, and natural language processing techniques described, before these are used in a small-scale exploration of clustering and dimensionality reduction techniques to find metaphors, and their associated interpretative repertoires. A larger-scale topic modelling analysis then explores the interpretative repertoires and metaphoric language of cancer.net in more detail, and a dictionary of potentially metaphoric terms is compiled. Finally, word vector similarity is investigated as a means of locating other metaphor in the data.

From this point the names of linguistic variables, including identified metaphor themes, and metaphoric terms and phrases, are italicised.

2.2 Cancer and metaphor

Sontag (1978) describes the specific illnesses cancer and tuberculosis as having been "spectacularly, and similarly, encumbered by the trappings of metaphor" (Sontag, 1978, p. 5) both in terms of the metaphors used to frame them, and the use of the illness itself as

a metaphoric framing of other experience. For example, a recent systematic analysis found that *cancer* may be used as a metaphor "to legitimize extreme measures as solutions to perceived threats" (Potts and Semino, 2019, p. 94), a use which is insensitive to, and unhelpful for, those living with cancer. In terms of metaphors used to address the experience of living with illness more generally, the master metaphor list (Lakoff, Espenson and Schwartz, 1991) describes the pervasive conceptual metaphor TREATING ILLNESS IS FIGHTING A WAR, in which, for example, 'The Disease is an Enemy', and 'The Body is a Battleground'. This is reflected in the use of metaphor by UK cancer charities which, apart from Maggie's and Marie Curie, use a confrontation (fight) metaphor (underlined in the following table) as their prominent website message:

Table 2-1 UK cancer charities fight-based slogans

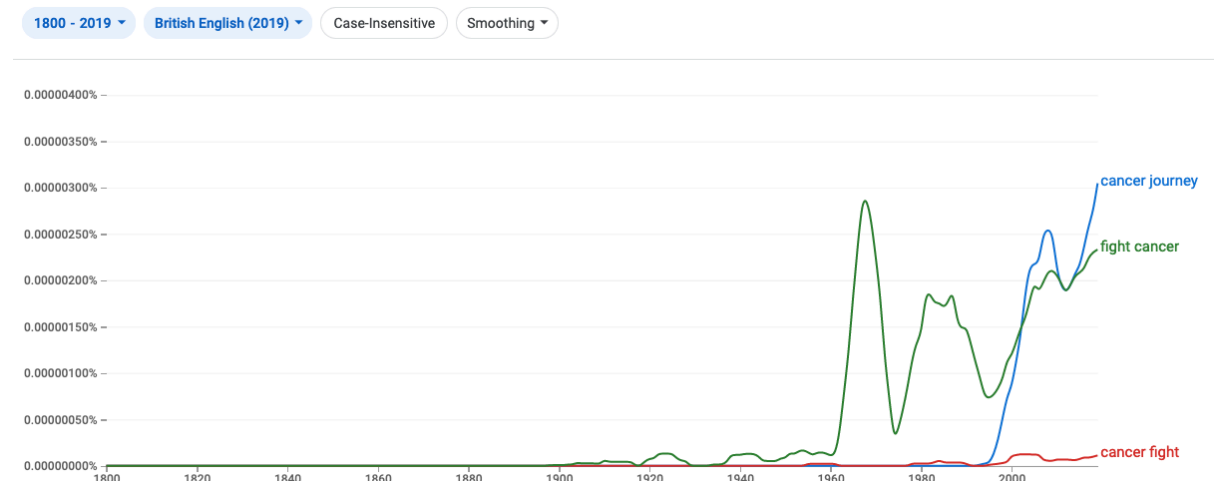
charity	message
Cancer research UK	Together we will <u>beat</u> cancer Stand Up to Cancer
Macmillan	Whatever cancer <u>throws</u> your way, we're right there with you
Childhood Cancer Parents Alliance	United in helping families <u>fight</u> childhood and adolescent cancer across the UK
Click Sargent	Young Lives <u>vs</u> Cancer

A large scale corpus study of metaphor in relation to cancer and the end of life (MCEL) found violence and journey metaphors to be the two most frequent types of metaphor across each of the six sections of their corpus, with violence the most used metaphor type for each participant group investigated, each of which "use a much wider set of violence metaphors than has been noted in previous studies, and each group uses them in a range of ways for different topics" (Semino *et al.*, 2017, p. 99). It is also noted in the study that violence and journey metaphors are often used together.

Entering into the Google Books n-gram Viewer the dominant metaphoric terms for discussion of cancer, *cancer journey*, *cancer fight*, and *fight cancer*, shows that the

framing of cancer as a journey was little used before the 1990s, but rose suddenly to become the most dominant of these metaphors by the early 2000s.

Figure 2-1 use of journey and fight metaphors for cancer over time, Google Books n-gram Viewer, British English



This raises the question whether cancer charities are out of step with current thinking around this issue. For example, an experimental study found the belief that a patient is likely to feel more guilty if they do not recover from cancer if their experience is framed as a battle than if it is framed as a journey (Hendricks *et al.*, 2018); another study found that "bellicose metaphors for cancer can influence the health beliefs of nonpatients in ways that may make them less willing to enact healthy behaviors" (Hauser & Schwarz, 2020, p. 1698). Another study found that the fight metaphor may be used by cancer patients as a way to protect others by concealing emotional distress, and that where healthcare professionals encourage patients to 'fight', and to have a positive attitude, they effectively promote and collude with emotional suppression of this difficult experience (Byrne A *et al.*, 2002). However, as discussed in chapter one, the MCEL study found that while some people may feel disempowered by violence metaphors, others find them motivational. In comparison while the journey framing does not tend to lead to feelings of failure and may generally be seen as a positive approach to the experience of cancer, for

some people it represents a sense of not being in control, and of frustration. The Cancer Research UK charity similarly notes that while violence metaphors may be unhelpful in the context of the individual experience of cancer, when it comes to fundraising for cancer they can be unifying and motivational (Gajewski, 2015).

Since journey and fight metaphors are prevalent in the discussion of cancer, they are likely to be prevalent in the cancer.net data, and to not, in a general way, be part of its specific language identity. However, consideration in this investigation is also given to the specific language used to express these, and other, metaphors in this context, and to the specific social and psychological work they do within cancer.net. In terms of other metaphors, the 'metaphor menu' (Semino, 2019) established as a result of the MCEL study (Semino *et al.*, 2017), suggests that the 'Fairground ride', 'Stone in your shoe', 'Music', 'Invasion', 'New Value', and 'Nature', metaphoric imagery found in the study may be useful to different people, and at different times in the experience of illness. Other studies have found that imagery of natural disasters (Bowker, 1996), personification (Gibbs and Franks, 2002), and plants and animals (Harrow *et al.*, 2008), reveal a pervasive understanding of cancer as a living entity.

2.3 The cancer.net data

2.3.1 Ethics

This project has received ethical approval from the University of Birmingham: ERN_20-0474.

In considering ethical issues relating to the current research for both cancer.net and suicide.net, which will be considered in chapter six, the good practice guidelines of the British Association of Applied Linguistics (BAAL, 2021), The British Psychological

Society ethics guidelines for internet-mediated research (BPS, 2017), and the ethical guidelines of the Association of Internet Researchers (AoIR, 2019) were consulted.

The British Association for Applied Linguistics (BAAL) guidelines for good practice state that it is acceptable from an ethics viewpoint to use linguistic frequency information from internet discourse, which is the approach of the current study, without specific consent. The Association of Internet Researchers ethical guidelines state that gaining consent is impractical for big data projects such as the current study. While some researchers do try to obtain informed consent, others focus on deleting names and other highly identifiable information when storing and processing data, which is the approach taken in the current study. And while consent might be obtained for the publication of a quote or other data that might make someone identifiable, the guidelines state that this may be unnecessarily prohibitive in terms of the practical effect on the creators of such data (AoIR, 2019).

The British Psychological Society ethics guidelines for internet-mediated research state that consent is not necessary where (i) online data can be considered to be in the public domain, and (ii) undisclosed use is justifiable on the grounds of scientific value (BPS, 2017). The usage guidelines of both the fora investigated in the current study state that the data created by participants is in the public domain. In addition, it is the intention that valuable insights about the experience of the health conditions investigated will be created from the current research.

2.3.1.1 Studies used as the ethical model for the current research

In addition to considering established ethical guidelines, two recent corpus linguistic studies were used as a model for the ethical approach of the current investigation. In contrast to the current study these two model studies undertake a more qualitative

analysis and make more quotes from the fora they investigate. It was reasoned that if their ethical approach is followed, it should be more than adequate for the current study, which is predominantly based on linguistic frequencies.

Like the current study, the first study used as a model for ethics, *Metaphor, Cancer, and the End of Life (MCEL)* (Semino *et al.*, 2017), uses data from the forum of a UK cancer charity. The MCEL approach is that there is arguably a consensus about the ethics of using online data, such that anything posted in public can be used as research material without the need to seek informed consent from individual participants, as long as anonymity is preserved (Semino *et al.*, 2017), which is consistent with the ethical guidelines discussed above. To confirm this approach the professional administrators of cancer.net were contacted for permission to use cancer.net data in the current research; they replied "All content on [cancer.net] is viewable publicly and therefore we are happy to allow you to obtain the data required for your research."

With regards to the participant-led forum suicide.net, which is considered in chapter six, the situation is different. The comparative research used as a model for ethics for suicide.net is *Corpus Discourse and Mental Health* (Hunt and Brookes, 2020), which analyses data from three participant-led fora associated with the mental health issues of depression, anorexia, and the contested condition diabulimia. Their in-depth consideration of ethics led to the conclusion, made in reference to BAAL guidelines, that permission should not be sought from these participant-led communities because the greater risk is to disturb them. Requesting informed consent risks disrupting the fora and undermining their primary purpose, such that "We perceived this risk to be greater than the potential harm that covert research would pose to contributors" (Hunt and Brookes, 2020, p. 81).

2.3.2 Collecting data from cancer.net

The Scrapy Python framework (Scrapy, 2021) was used to write web 'spiders' to collect data from the cancer.net forum. Data was not collected for participants formally associated with the forum, for example moderators and health care professionals, since the focus of the current study is on the experience of living with cancer for patients, or friends or relatives of cancer patients. Anonymous posts were also excluded for all analysis based on the number of posts made since the composite 'anonymous' participant would distort such analysis. Data from the composite anonymous participant is retained for all other analysis. It should be noted that participants may not be anonymous through routine choice: this status is assigned by moderators on the request of participants and their families or friends, perhaps when a participant has died. For that reason, there is no analytic interest in posts from the composite anonymous participant specifically: there is no reason they should be notably different from non-anonymous posts.

Cancer.net is a very active and diverse forum that had been running for over 12 years at the date of data collection. Its current state reflects many historical and ongoing interactions and requirements. It was not possible to collect every single post ever made, since posts and threads may be deleted, with sometimes the whole history for a particular participant removed from the forum. Sometimes, as discussed above, rather than being removed posts for particular participants are designated 'anonymous'. However, all posts were collected that were still available at the time of collection. Every thread on the forum was crawled to pick up all posts on that thread, with each post saved as a separate record in a single JSON file, where JSON stands for JavaScript Object Notation, a common and useful data interchange format. The following data was collected for each post:

1. the id of the participant who made the post (anonymised)
2. the title of the thread the post was taken from
3. the id of the thread the post was taken from
4. the date the post was made
5. the text content of the post

2.3.3 Processing the data

The JSON data file was then processed to remove posts outside the focus date range (posts from 01/03/2021 onwards), and the remaining posts were sorted into ascending date order and converted to lower case.

Because of the way the cancer.net website works, each post from the forum is returned as a list. The list was filtered so that empty items were removed, and if the subsequent length of the list was greater than 2, the initial and final list items, which are typically greetings and farewells, were removed.

The data was cleaned to remove formatting codes that are present in the text data as a result of the way the forum works. These are listed in the table below.

Table 2-2 formatting codes removed from cancer.net post text

code (removed)	represents
<code>\n</code>	new line
<code>\u00a0</code>	non-breaking space

The processed data was 'pickled' as a Python Pandas DataFrame, where a DataFrame is a powerful way of storing data that allows it to be queried as a relational database, and to be analysed using the many DataFrame data analysis methods. To ensure consistency, each subsequent analysis uses this same stored data.

2.3.4 Participants and posts: statistics and distribution over time

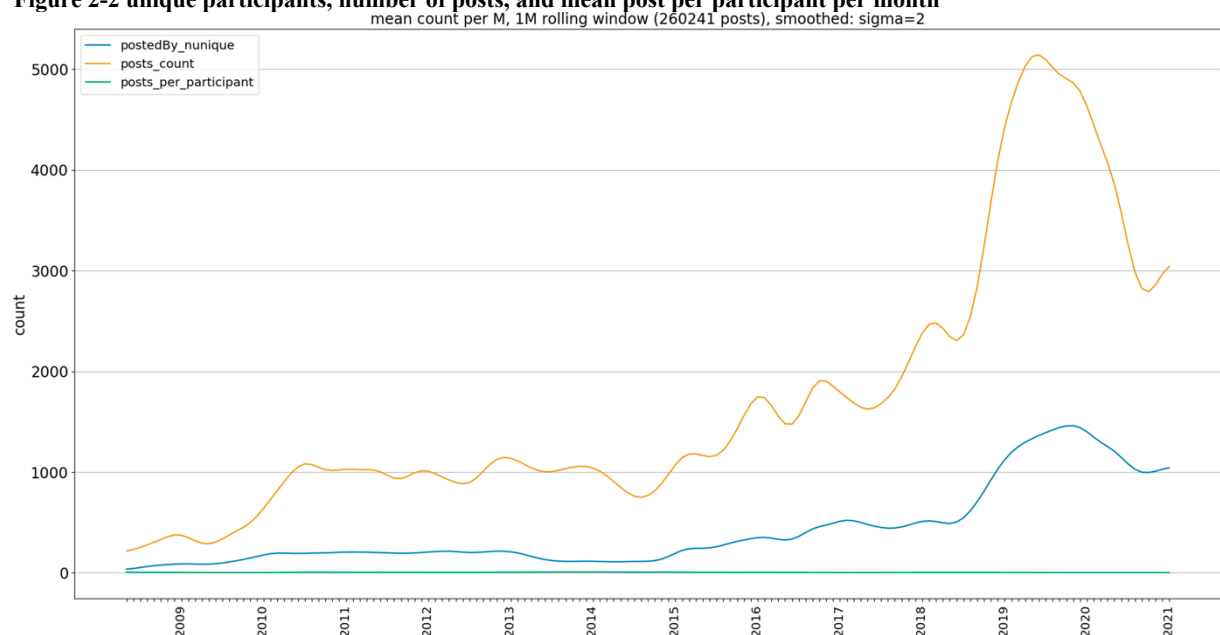
The cancer.net data consists of all posts from the start of the forum, 06/06/2008, up to the last post made on 28/02/2021, with data collected on 07/03/2021. The following table summarises relevant statistics for this data.

Table 2-3 cancer.net statistics

count	description	referred to in later analysis as
260,239	posts collected	
42,311	participants with at least 1 post	
21,265	participants with 1 post	
41,021	participants with < 20 posts	cancer.net.LT20
1,290	participants with >19 posts	cancer.net.20
20	participants with > 1,000 posts	
6	average number of posts per active participant	

The following chart shows the number of unique participants (blue), number of posts (orange), and mean number of posts per unique participant (green), per month, since the forum started. The mean number of posts per active participant per month varies between a minimum of 2 and a maximum of 11.

Figure 2-2 unique participants, number of posts, and mean post per participant per month

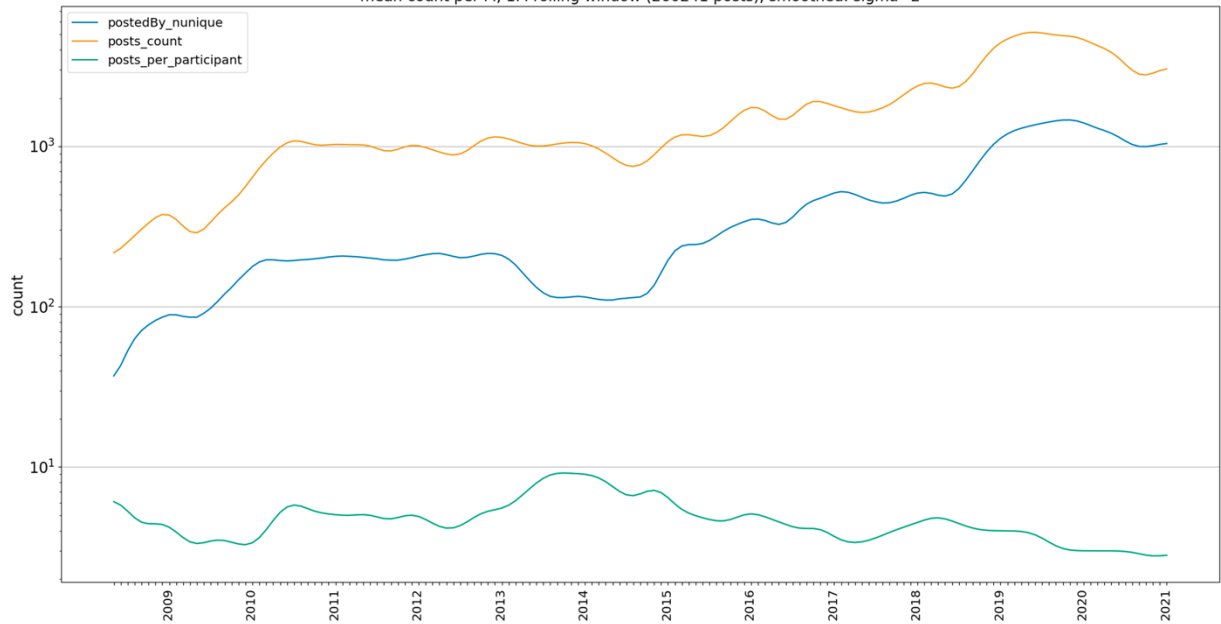


The same data plotted against a log y axis, below, shows more clearly that the number of unique participants and number of posts made in a month track each other, while the mean number of posts per active participant per month shows the inverse behaviour. In general, then, an increase in unique participants in a month reduces the mean number of posts each participant makes in a month, even though a greater number of unique participants in a month means that there are more posts in that month. This suggests a core of regular participants who post relatively frequently, while participants who are new to the forum make relatively fewer posts. This is supported by a correlation analysis of these variables by month. The table below shows that while the number of posts per month is strongly correlated with number of unique participants per month ($r=0.967609$), the number of posts per participant has a weak (close to moderate) negative correlation with number of unique participants per month ($r=-0.481176$), and the number of posts per participant similarly has a weak negative correlation with number of posts per month ($r=-0.33476$).

Table 2-4 correlation of variables by month

	posts	unique_participants	posts_per_participant
posts	1.000000	0.967609	-0.333476
unique_participants	0.967609	1.000000	-0.481176
posts_per_participant	-0.333476	-0.481176	1.000000

Figure 2-3 unique participants, number of posts, and mean post per participant per month, logy axis
 mean count per M, 1M rolling window (260241 posts), smoothed: sigma=2



2.3.5 cancer.net as a community

The distribution of participant posts on cancer.net, discussed above, meets the participation criterion of an online community, discussed in chapter one, in that it demonstrates "Frequent, regular, self-sustaining activity over time" (Herring, 2004, p. 361). In addition, there are a number of extended threads on cancer.net that have been active over multiple years. For example, a general thread encouraging participants to consider the good and the bad in their current experience was started by a participant in 2017, and was still active, with over 6,000 posts in total, at the end of February 2021. In conjunction with distribution of posts by participant, this meets the interaction criterion of an online community, which requires "Reciprocity, extended (in-depth) threads, core participants" (Herring, 2004, p. 361). Although more criteria of community are considered in future analysis, from this point on cancer.net is understood as a community.

2.3.6 Section summary: the cancer.net data

Cancer.net is a large and diverse set of naturally occurring text data spanning more than 12 years. All available posts from the cancer.net forum were collected for analysis. The pattern of participation on cancer.net meets two of the criteria for a virtual community.

2.4 Keywords and key terms

For the analysis in this section the Sketch Engine corpus analysis website was used to calculate keywords and key terms for cancer.net, where the keyness of terms, which is typically used to explore what a corpus is 'about', relates to their frequency in the corpus of focus compared to their frequency in a reference corpus, which in this investigation is the general web corpus enTenTen18.

In the following results tables the keywords and key terms are ranked in descending order of their score, which is a measure of the significance of the difference between their density in cancer.net in comparison to the general web corpus. The score measure used is the simple maths parameter (SMP): the frequency per million of a term in the focus corpus, divided by the frequency per million of a term in the reference corpus, with a value N , for which the default is 1, added to each frequency value before division as a smoothing parameter. As the value of N is increased, so the focus is increasingly on less rare terms. As the value of N is decreased, so the focus is increasingly on more rare terms.

While log likelihood is often used for corpus comparison, there is growing evidence that it tends to identify too many keywords, some of which may be regarded as false positives (Brezina, 2018). To address this Kilgariff (2009) suggests consideration of the ratio between relative frequencies of words in comparative corpora. To address the related issue that such a ratio can only be calculated if values are greater than zero, Kilgariff

suggests adding a constant to relative frequencies before calculating the ratio. This formula is the simple maths parameter. Through adjustment of the SMP smoothing parameter, described above, this method also allows focus on terms having different levels of rarity, which is productive in the following analysis.

2.4.1 Keywords for cancer.net in comparison with a general web corpus

For the analysis in this section cancer.net was compared with the general web corpus English Web 2018 (enTenTen18) in order to locate its keywords. When the score parameter N was set to 1, so that more the focus is on more rare terms, the keywords, listed on the left in the table below, mostly contain references to cancer and its treatment. There are also references to Christmas, as the keyword lemmas *christma* and *xma*, which is an important reference point or milestone for cancer.net participants, as well as representing the good things about life, for example when families meet up and celebrate. The other notable set of keywords relates to older generation relatives, including *mum*, *hubby*, *grandad*, and *dad*, although these may relate to language differences between the UK-based cancer.net corpus and the general web corpus.

When the score parameter N was set to 1,000, so that the focus is on more common terms, on the right in the table below, the most significant keywords are the first-person singular pronouns *i*, and *my*, while *me* is also a notable keyword (marked with asterisks). Older generation references include *mum*, and *dad*, while the personal pronoun *she*, referring to the other, is also a keyword. The greater prevalence of first-person singular pronouns in cancer.net, for example *i* is used over six times as often in comparison to the reference corpus, is suggestive of worse mental health on cancer.net in comparison to general web participation. There are no potentially metaphoric keywords in the output from either analysis.

Table 2-5 keywords, score parameter N=1 (left) and N=1,000 (right)

keywords N=1	per M (focus)	per M (ref)	score	keywords N=1,000	per M (focus)	per M (ref)	score
1. chemo	1416.20	2.3	432.0	*i	35709.4	5649.9	5.5
2. mum	2057.4	8.8	210.2	*my	12797.4	1579.3	5.3
3. radiotherapy	451.7	2.4	131.7	cancer	4033.4	131.2	4.4
4. christma	272.3	1.3	118.2	feel	3817.4	399.10	3.4
5. hubby	314.4	2.3	95.7	*me	5210.0	864.1	3.3
6. lumpectomy	119.5	0.3	91.5	so	7728.7	1697.2	3.2
7. oncologist	278.9	2.1	90.6	mum	2057.4	8.8	3.0
8. covid	86.8	0.0	87.6	go	5535.8	1220.0	2.9
9. mastectomy	172.3	1.0	87.5	hope	2492.9	204.9	2.9
10. biopsy	579.4	6.9	73.4	you	16674.9	5190.8	2.9
11. lump	618.1	7.7	70.8	get	5517.7	1403.0	2.7
12. lymph	394.7	5.2	63.6	just	4523.5	1048.4	2.7
13. gp	683.4	10.3	60.8	have	23366.2	8311.2	2.6
14. grandad	79.2	0.4	58.2	week	2491.7	349.3	2.6
15. tumour	386.1	5.9	56.5	she	4349.7	1070.1	2.6
16. mammogram	124.5	1.3	53.5	know	4088.4	1008.7	2.5
17. xma	64.3	0.3	49.6	chemo	1416.2	2.3	2.4
18. dad	1,443	28.4	49	think	3108.8	720.3	2.4
19. tamoxifen	89.8	1	45.3	dad	1443.0	28.4	2.4
20. letrozole	52.2	0.2	44.4	hi	1455.0	36.3	2.4

2.4.2 Key terms for cancer.net in comparison with a general web corpus

The key terms (multi-word expressions) for cancer.net in comparison with the general web corpus, listed in the table below, provide further insight into the specific concerns of the forum. When the score parameter N was set to 1, on the left of the table, so that the focus is on more rare terms, the metaphoric term *cancer journey* is present (underlined). Otherwise, the key terms are similar when N was set to 1 (on the left of the table) to focus on more rare terms, and when it was set to 1,000 (on the right of the table) to focus on more common terms. They include specific terms relating to cancer, such as *ct scan* and *breast cancer*, and greetings addressed to the other, such as *much love*, and *good luck*.

Table 2-6 key terms, score parameter N=1 (left) and N=1,000 (right)

key terms N=1	per M (focus)	per M (ref)	Score	key terms N=1,000	per M (focus)	per M (ref)	Score
1. ct scan	163.0	0.0	159.4	breast cancer	411.0	13.2	1.4
2. bowel cancer	116.1	0.3	93.5	good luck	333.5	9.4	1.3
3. breast clinic	89.9	0.0	89.7	lung cancer	217.0	4.8	1.2
4. pet scan	55.8	0.2	47.1	next week	191.0	10.5	1.2
5. lymph node	106.1	1.5	42.2	ct scan	163.0	0.0	1.2
6. lung cancer	217.0	4.8	37.6	good news	171.7	12.1	1.2
7. cancer chat	36.2	0.0	37.2	last week	175.0	30.7	1.1
8. right breast	41.7	0.2	35.8	last year	185.9	56.7	1.1
9. mri scan	33.7	0.0	34.6	bowel cancer	116.1	0.3	1.1
10. much love	53.4	0.7	33.0	lymph node	106.1	1.5	1.1
11. big hug	38.5	0.2	32.5	breast clinic	89.9	0.0	1.1
12. double mastectomy	34.0	0.1	32.5	blood test	73.8	1.6	1.1
13. good luck	333.5	9.4	32.2	last night	87.3	15.1	1.1
14. hope everything	30.5	0.1	29.8	prostate cancer	70.3	5.1	1.1
15. <u>cancer journey</u>	30.6	0.1	29.4	long time	93.8	29.8	1.1
16. breast cancer	411.0	13.2	29.1	treatment plan	60.0	1.6	1.1
17. macmillan nurse	28.0	0.0	29.0	pet scan	55.8	0.2	1.1
18. blood test	73.8	1.6	28.7	much love	53.4	0.7	1.1
19. brain tumour	33.9	0.2	28.5	bad news	56.6	3.9	1.1
20. terminal cancer	32.2	0.2	27.9	new year	55.2	3.4	1.1

2.4.3 Section summary: keywords and key terms

Analysis of keywords and key terms for cancer.net in comparison to a general web corpus showed its primary concerns to be specifically related to cancer, while family references were also found to be relatively prevalent. The single metaphoric term *cancer journey* was present as a key term, suggesting that it is a dominant metaphor of this corpus, in which the focus is on metaphor instantiated as nouns, and potentially a metaphoreme. A focus on more common terms showed that there is a much higher prevalence of first-person singular pronouns in cancer.net in comparison to the general web corpus, a language style associated with worse mental health. This is a key focus of the current investigation and is considered in more detail in later analysis.

2.5 Selecting data based on the number of posts made

In order to find typical metaphors of the cancer.net community, it is necessary to identify what part of the data best represents that community. The many participants who have made only one post, for example, are unlikely to truly represent the discourse of cancer.net. To investigate this issue, the Scattertext Python library, "a tool for visualizing linguistic variation between document categories" (Kessler, 2017), was used to compare data based on the number of posts made by participants - to compare data between groups of participants who are more, or less, established within the community. The categories used for this broad-brush exploratory comparison were participants who have made more than 1,000 posts, compared with participants who have made fewer than 20 posts, with 20 selected as a balance between considering participants who have been relatively active on and involved with the forum (half the participants have only ever made one post, for example), while still having made a relatively small number of posts (1,290 participants have made 20 or more posts).

Running Scattertext on the 500 most used terms for the comparative participant groups showed that the newer forum participants use more literal and specific language to focus on their current personal physical lived experience, while established forum participants use more figurative language. Terms that are potentially metaphoric in this context are underlined in the following summary tables.

Table 2-7 top terms for participants with a post count <20

term type	top terms for participants with a post count <20
tokens	i'm, neck, tumour, bowel, blood, removed, stomach, ultrasound, ct, biopsy, ok
phrases	5 weeks, 3 weeks, 6 weeks, blood test, left side, right side, left breast

Table 2-8 top terms for participants with a post count >1,000

term type	top terms for participants with a post count > 1,000
tokens	welcome, hubby, ok, bouts, we're, <u>journey</u> , forum, hold, feelings, <u>ring</u> , psa, number, double
phrases	2 bouts, bouts of breast cancer, <u>cancer journey</u> , warm welcome, <u>boxing gloves</u>

The group of participants who have made more than 1,000 posts made 24 uses of the potentially metaphoric term *journey* per 25,000 terms, and 94 uses per 1,000 posts. In comparison, the group of participants who have made fewer than 20 posts made 4 uses of *journey* per 25,000 terms, and 16 uses per 1,000 posts.

In order to make the majority of terms legible for the following static screenshot of the interactive Scattertext visualisation, just 100 top terms were included, with terms included only if they occur at least five times. In the figure "each axis corresponds to the rank frequency a term occurs in a category of documents" (Kessler, 2017). In the upper-left corner of the plot are terms frequently used by participants with more than 1,000 posts but infrequently or never used by participants with fewer than 20 posts. In the bottom-right corner are terms frequently used by participants with fewer than 20 posts, terms that are infrequently or never used by participants with more than 1,000 posts.

Since the purpose of the current chapter is to find metaphor themes that are characteristic of the cancer.net community, the data used for subsequent exploratory analyses in this chapter excludes participants who have made fewer than 20 posts, whose language understandably has an overwhelmingly literal focus on their own current experience. At the time the data was collected 1,290 participants had made 20 or more posts, accounting for 141,424 posts in total. This subset of cancer.net will henceforth be referred to as

cancer.net.20. The MCEL project discussed above similarly uses online data only for those who have written at least 1,000 words, because "the discourse of an online forum is better typified by the linguistic behaviour of habitual users than by one-off contributions from passers-by" (Semino *et al.*, 2017, p. 57).

2.5.1 Section summary: selecting data based on number of posts made

It was found that cancer.net participants who have made fewer than 20 posts tend to use more literal language in comparison to participants who have made more than 1,000 posts. Since participants who have made fewer than 20 posts are also not such an established part of the discourse community, the initial exploratory analyses to prospect for metaphor in cancer.net use data from participants who have made 20 or more posts. This subset of data consisting of 141,424 posts from 1,290 participants will from now on be referred to as cancer.net.20.

In the next section an initial exploration is made of the use of unsupervised learning techniques to find metaphor in cancer.net.

2.6 Unsupervised learning introduction

2.6.1 Automatic metaphor identification

Identifying metaphor automatically in text is recognised as a challenging problem that is essential for a number of applied natural language processing tasks, including machine translation; opinion mining; and information retrieval (Shutova, 2015). It is also critical to the pressing need for "greater precision in reliably identifying conceptual metaphors from the systematic analysis of language patterns" (Gibbs, 2017, p. 265). The process of trying to automate metaphor identification also offers interesting insights into how metaphor works, with identification systems typically applying very specific rules that may surpass

humans in identifying specific types of metaphor (Gibbs, 2017). But although there is much current interest in the automatic identification of metaphor, it remains an unsolved task (Veale, Shutova and Klebanov, 2016).

A report on the Association for Computational Linguistics 2020 metaphor detection task notes that over the last decade approaches to automatic metaphor detection have explored features based on "concreteness and imageability, semantic classification using WordNet, FrameNet, VerbNet, SUMO ontology, property norms, and distributional semantic models, syntactic dependency patterns, sensorial and vision-based features" (Leong *et al.*, 2020, p. 1). But it has been suggested that hand-constructed resources such as WordNet do not have the density and diversity of connections needed to support realistic automatic interpretation of metaphor (Veale, Shutova and Klebanov, 2016). More recently focus has switched to the use of deep learning models, with more than half the entries for the Association for Computational Linguistics 2020 metaphor detection task using BERT (Bidirectional Encoder Representations from Transformers), a transformer-based deep learning model trained on unlabelled text, which considers left and right contexts simultaneously, across all model layers (Devlin *et al.*, 2019).

In a large-scale systematic study of metaphors, cancer, and the end of life (MCEL) (Semino *et al.*, 2017), qualitative and quantitative corpus-based analysis were combined to identify patterns of metaphor in a six-part 1.5 million word corpus representing (i) patients with advanced cancer, (ii) unpaid family carers, and (iii) healthcare professionals, with data for each from (a) semi-structured interviews, and (b) naturally occurring online forum posts. Manual identification of metaphor in a structured 92,000 word sample was used to support both lexical and semantic concordances in the whole corpus. Metaphors were found in the following USAS (UCREL Semantic Analysis System) (Rayson *et al.*,

2004) semantic categories: Animals, Journey, Machines, Obstacles, Openness, Religion and the Supernatural, Restraint, Sports and Games, Violence, and Wholeness. Wmatrix corpus linguistics software was then used to identify all data annotated with semantic tags relating to these categories, although because there are many literal journey references in the data this method was not found to be productive for the Journey group. In addition, searches were made for particular formations that may denote use of metaphor or simile, for example *like* as a preposition, something that is explored in chapters three and six of the current study in investigations of signalled metaphor.

In order to remain open to what metaphor is present in the diverse, naturally-occurring data under consideration here - to undertake a corpus-driven rather than a corpus-based investigation (Deignan, 2005) - the investigations in the rest of this chapter use unsupervised learning techniques to prospect for metaphors in the data. In unsupervised learning, transformations of unlabelled data create new representations of that data that may provide new insights, which may then be used to support analysis using other methods. Unsupervised learning consists of all machine learning for which it is not already known what 'answers' are sought (Müller and Guido, 2016).

Machine learning is about extracting knowledge from data (Müller and Guido, 2016, p. 1)

The following section addresses some of the pre-processing, and other analytic considerations, that are required to be able to use unsupervised learning techniques with text data such as cancer.net.

2.6.2 Natural language processing

In the current investigation text data was processed using the spaCy natural language processing Python library (Honnibal and Montani, 2021). Using statistical language models to process text according to its context, the spaCy parser is quick and has high accuracy, which is evaluated rigorously. Its English part of speech tagger uses the OntoNotes5 version of the Penn Treebank tag set. In addition, tags are mapped to the Universal Dependencies v2 POS tag set. As well as being accurate and consistent, automatic tokenization and tagging can avoid disagreements over the manual identification of lexical units prior to the identification of metaphor (Veale, Shutova and Klebanov, 2016).

The spaCy transformer-based language model *en_core_web_trf* and non-transformer-based language model *en_core_web_lg* are highly accurate web-based models trained on written text from the internet (blogs, news, and comments). The following table details the accuracy of these two language models for different methods:

Table 2-9 accuracy of spaCy language models

		en_core_web_trf	en_core_web_lg
TOKEN_ACC	Tokenization	1.00	1.00
TAG_ACC	Part-of-speech tags (fine grained tags, Token.tag)	0.98	0.97
DEP_UAS	Unlabelled dependencies	0.95	0.92
DEP_LAS	Labelled dependencies	0.94	0.90
ENTS_F	Named entities (F-score)	0.90	0.86
SENTS_F	Sentence segmentation (F-score)	0.89	0.89

Although the spaCy transformer-based model *en_core_web_trf* is more accurate, it is also very resource-heavy, slower, and places limitations on segment size of text. For this reason, the still highly accurate *en-core_web_lg* language model was used in the current investigation.

2.6.3 Lexical unit of focus

For some analyses it is more productive to focus on a subset of tokens, or to focus on a particular lexical unit. In the current metaphor-focused investigation, since many metaphors take the form of multi-word expressions rather than a single word (Cameron *et al.*, 2009; Low *et al.*, 2010), as well as being considered as individual tokens, text is also organised into n-grams and noun chunks for analysis, as discussed below.

2.6.3.1 Individual tokens

In the current study individual tokens are identified using the spaCy tokenizer. Since spaCy can identify parts of speech, it may be used to select only the nouns present in the data, for example, or the lemmas of those nouns, while for some analyses all parts of speech may be used.

2.6.3.2 n-grams

For some investigations in the current study n-grams are sought, where n-grams are the collocations (common phrases) of a specified length of the lexical terms of focus. For example, bigrams are collocations of two lexical terms; trigrams are collocations of three lexical terms. In the current investigation n-grams are processed using the Gensim Python library *Phraser*, which automatically detects n-gram collocations from a stream of sentences.

2.6.3.3 Noun chunks

For some analyses in the current study spaCy is used to generate noun chunks from the data, where a noun chunk is a base noun phrase - a flat phrase that has a noun as its head. "You can think of noun chunks as a noun plus the words describing the noun" (Honribal

and Montani, 2021). A spaCy noun chunk will be referred to as a noun phrase from now on, to avoid confusion over the introduction of a term that is not typical in linguistics.

2.6.3.4 Focusing on nouns in order to find metaphoric language

Nouns refer to things that have spatial dimensions. Because of this nouns may make any unconventional (potentially metaphoric) reference more accessible: they may be the most recognizable and vivid of metaphors (Goatly, 2011). The Metaphor Identification Procedure (Pragglejaz, 2007) similarly notes that it is generally easiest to assign a basic sense to nouns, making the understanding of metaphoric nouns as metaphoric more accessible compared to metaphor based on other parts of speech. And since noun-based metaphor has more potential to activate other sensory metaphoric associations, according to the dynamic view it is potentially more likely to activate metaphorical thinking (Müller, 2008). The focus in the current study on nouns in individual phrases, sentences, and posts, is at a fundamental level not consistent with the discourse dynamics framework. This issue is addressed in more detail at the end of this thesis in section 7.2.1, in terms of the relevance of the discourse dynamics approach to dynamic considerations in the thesis. Those dynamic considerations are explored as the diachronic analyses of chapter 5 and section 6.9, which are based on counts of metaphor instantiated as nouns.

Goatly (2011) suggests the following hierarchy of force of metaphor source terms:

1. nouns (most active)
2. verbs and adjectives (less active)
3. adverbs and prepositions (least active)

As the current investigation is concerned with the measurable psychological impact of metaphor use the primary focus is on finding metaphors for which the source term is a noun.

The viability of this approach for the purposes of the current investigation is considered in more detail in the supervised machine learning analysis of chapter four. In chapter four, machine learning models are trained to predict the presence of the identified metaphor themes in each post in the data. For this purpose machine learning algorithms are supplied with a set of training posts labelled as containing or not containing each identified theme, based on the presence or absence in each post of terms from that theme instantiated as nouns, with identified metaphoric terms first removed from the training data supplied to the algorithm. If the models accurately predict the presence of each theme in any post based on the language style of the whole post, this supports the focus on nouns for the purposes of the current investigation. To further explore this issue, in chapter four a comparison is made between identified metaphoric terms instantiated as nouns only, and as nouns and verbs.

2.6.4 Vectorization

The data science methods used in the current investigation typically work with numeric tensors, not text. In order to use text with these methods, once the lexical items of focus (terms) are extracted from the text data, for example all the noun lemmas, those terms are then vectorized. Vectorization is the process of turning a collection of text-based documents into numerical feature vectors called a 'bag of words' representation, where 'words' represents the lexical unit of focus. After vectorization, the text documents are described by occurrences of the terms contained in them, while the relative position of those terms in the document is ignored. That loss of information about relative position is

why generating n-grams or noun phrases prior to vectorization can be more revealing of information about the text for some analyses, since those methods retain some information about word order.

As well as creating the vector of numerical features representing the terms in the text data, the vectorization process also creates a dictionary detailing which text term is associated with which numerical feature. So, when the analysis requiring numerical tensors has been run, the output can be converted back to text for display and interpretation. The scikit-learn Python library *CountVectorizer* (Pedregosa *et al.*, 2018) is used in the current investigation to vectorize text.

Different methods for working with text data are used in the current study for different analytic purposes. In the current chapter the text data is vectorized for use with unsupervised learning methods, to look for metaphoric terms in the data. As described above this 'bag of words' approach does not take into account the order of terms in a document. It is the use of terms together in the same post or to discuss the same topics that is the focus of this first analysis: the order of terms within the post is not important.

The final section in the current chapter in comparison uses word vectors to further explore potentially metaphoric language by locating terms in the data that have semantic similarity to potentially metaphoric terms found via unsupervised methods. In comparison to the bag of words approach, word vectors encode semantic relationships between words, based on general usage (Honnibal & Montani, 2021). In chapter four, two comparative supervised learning methods are used to predict the presence of metaphor in posts. One of these, a text-based method, supplies whole texts to a supervised algorithm, an analysis which does take word order into account. The other supervised method counts

the presence in each post of 143 different linguistic variables, for example nouns, the first-person singular pronoun I, and the number of polysyllabic words.

2.6.5 Term frequency

The `max_df` (maximum document frequency) and `min_df` (minimum document frequency) parameters supplied to the *CountVectorizer* facilitate a finely grained exploration of the data based on the prevalence of terms. If a floating-point number is supplied in the range 0.0 to 1.0 this refers to the proportion of documents in which a term is present. If an integer is supplied, this refers to the number of documents in which a term is present. Changing these values can radically change the output.

Table 2-10 definitions of `max_df` and `min_df`

<code>max_df</code>	ignore terms that appear in more than this proportion or number of documents
<code>min_df</code>	ignore terms that appear in fewer than this proportion or number of documents

2.6.5.1 Controlling term frequency to look for metaphors

In setting the `max_df` value (the maximum documents in which a term may appear) the diversity of the data is considered. In looking for metaphors, the `max_df` value might be set higher if just one participant is being investigated, in comparison to all participants, because that one participant might use the same metaphoric language frequently. The choice of `max_df` value is also influenced by the lexical unit of focus. For example, if noun lemmas are the lexical unit of focus the `max_df` value might be set lower than if n-grams are the lexical unit of focus. Since n-grams are more complex and specific, each n-gram is likely to be more rare in the data than are many of the noun lemmas.

Since the investigative focus in the current chapter is on finding metaphor themes that are particularly representative of this language community, and not creative metaphors, the

min_df (the minimum documents in which a term may appear) is set above 1, so that terms are only considered that are present in more than one document in the corpus.

2.6.5.2 Comparison with keyness

The corpus linguistic technique of calculating the keyness of terms, explored above, compares their frequency in the corpus of focus against their frequency in a reference corpus of general language use, with differences between the two corpora being checked for statistical significance. Keyness is used to explore what a corpus is 'about', with terms that commonly occur alongside key terms acquiring keyness by association. But other lexical sets may be present in the low-frequency vocabulary of the focus corpus that do not obviously seem to be relevant to the 'aboutness' of the corpus and may therefore be missed by a focus on keyness. It is these groupings in low-frequency vocabulary that may suggest metaphorical activity (Phillip, 2010). Term frequency control can be used to locate those other meaningful terms within a corpus that do not have keyness - those terms that Phillip (2010) argues are more likely to be metaphorical.

2.6.6 Using k-means clustering to consider the effectiveness of term frequency control for finding metaphors

In this section a small k-means clustering investigation is used to consider the effectiveness of term frequency control for finding interpretative repertoires, and their associated metaphors, where clustering is the partitioning of data into groups, called clusters, consisting of data points (in this case linguistic terms) that typically occur together. K-means is a simple and commonly used unsupervised algorithm to locate cluster centres by which to group data based on similarity. The number of clusters to search for must be supplied to the algorithm and running the algorithm to find various numbers of clusters is another way of exploring the structure of the data. The simplicity

of k-means clustering can also be a weakness however: since each cluster is defined by its centre it is inevitably a simple shape, with each cluster having approximately the same diameter (Müller and Guido, 2016). K-means clustering is used for this initial investigation because it is a commonly used algorithm, and its simplicity supports the purpose of the investigation: if k-means clustering with term frequency control is found to be productive in finding interpretative repertoires and metaphors in the text, this supports term frequency control as a useful method for this purpose, which should be even more productive when used in conjunction with more sophisticated algorithms.

In k-means clustering, a random distribution of cluster centres is made, then the following two-step process is run until the assignment of data points to clusters no longer changes:

1. a data point is assigned to its nearest cluster centre
2. each cluster centre is calculated as the mean of each data point assigned to it

To consider the efficacy of using term frequency control to find metaphors and their associated interpretative repertoires in this large and diverse naturally occurring data, a small k-means clustering example was run to contrast output when term frequency is not controlled, with output when term frequency is controlled.

The scikit-learn Python library *Kmeans* algorithm was used to find three clusters, and in order to make the interpretation of clusters tractable only the 25 top terms were output. The data for these investigations consisted of noun lemmas, from which bigrams and trigrams were formed.

2.6.6.1 k-means clusters when term frequency is not controlled

In the following output, for which term frequency was not controlled, there are no potentially metaphoric terms in any of the clusters. In addition, there is much repetition of

terms between clusters, such that the dominant terms in the data exclude the rarer terms, as well as making a single cluster harder to consider as a separate interpretative repertoire. However, this is a good way of understanding what the dominant terms and themes are, which for cancer.net seem to be *cancer*, and *time*.

Table 2-11 results when term frequency is not controlled

cluster	cluster is about?	terms
1.	cancer and time	cancer year time treatment people thing week chemo day breast life way lot diagnosis month dad stage result family one doctor surgery hospital husband friend
2.	time and cancer	time day thing week year dad life cancer people way family lot bit chemo friend month treatment today one pain care work husband hospital daughter
3.	time and cancer	time day thing cancer week year treatment care people chemo today way lot bit result thank family one dad life news pain hospital friend luck

2.6.6.2 Determining the max_df value

Before using k-means clustering in conjunction with term frequency control, an exploration was made of suitable max_df values to use to focus the output on terms more likely to be metaphoric. Using cancer.net.20 in 10 trials of a random sampling of 50,000 posts, a maximum of 1,875 posts (3.75%), and a minimum of 1,783 posts (3.57%), were found to contain the potentially metaphoric token *journey* found in previous analyses in the current chapter. Since other metaphorical terms have not yet been found, it can be concluded that they are present in fewer documents than this, suggesting that it would be productive in terms of finding metaphors to set the max_df value to 0.05.

2.6.6.3 k-means clustering when term frequency is controlled

In order to focus on locating metaphors, for the following analysis, term frequency was controlled, with the max_df value set to 0.05 so that terms were only included that are

present in 5% or fewer of documents. Min_df was set to 2, so that terms were only included that are present in at least two documents.

Table 2-12 results when term frequency is controlled, with max_df=0.05 and min_df=2

cluster	cluster is about?	terms
1.	sending good wishes and advice	hug thought feeling tomorrow help site place problem hope advice wish end word heart <u>journey</u> situation person lady head experience mind nurse thread touch morning
2.	family, home, and time/event references	hubby daughter night home morning son tomorrow hour child weekend yesterday christmas house end bed couple weather garden place feeling wife course memory problem man
3.	medical care	surgery gp nurse stage consultant scan question radiotherapy problem test case biopsy oncologist patient symptom experience tumour advice surgeon lump lymph_node body help op site

When term frequency was controlled so that the focus is on less common terms the potentially metaphoric token *journey* is present in the output, in cluster one. Because there is less repetition of common terms between these three clusters, they are also easier to make sense of as distinct interpretative repertoires in comparison to the previous k-means example for which term frequency was not controlled. While cluster three is concerned with specific and literal aspects of treatment for cancer, and cluster two is concerned with practical arrangements around close family, cluster one, containing *journey*, appears to relate to general support and advice, and inclusion of the term *thread* in cluster one suggests that this cluster is more specifically related to the cancer.net forum.

2.6.7 Section summary: unsupervised learning introduction

In order to be used with quantitative methods that accept only numeric tensors, text must first be processed so that the lexical items of focus (for example all tokens, or all noun lemmas, or all noun phrases) are extracted from it. Those retained lexical items are then represented as a numerical tensor (vectorized), with the vectorization process also

providing a method for screening terms based on their frequency in the corpus, which allows, for the purposes of the current investigation, a finely-controlled focus on those more rare terms that may be more likely to be metaphoric. As the current investigation is concerned with the measurable psychological impact of metaphor use the primary focus is on finding metaphors for which the source term is a noun.

In the next section dimensionality reduction in conjunction with term frequency control is investigated as a method for finding potentially metaphoric language in the data, to explore the efficacy of different techniques.

2.7 Using dimensionality reduction to find interpretative repertoires

In the previous section the simple k-means clustering unsupervised algorithm in conjunction with term frequency control was found to be productive in finding interpretative repertoires and their associated metaphors in cancer.net. In the current section three dimensionality reduction unsupervised methods build on this technique, where dimensionality reduction is commonly used to reduce data to two or three dimensions in order to plot it; it is also used "to provide a lower-dimensional representation of documents that reflects concepts instead of raw terms" (Crain et al., 2012), which in the current investigation can be related to interpretative repertoires. The three dimensionality reduction methods Principal Component Analysis, factor analysis, and Latent Dirichlet Allocation, are explored with data consisting of noun lemmas from which bigrams and trigrams are formed. For each dimensionality reduction method, the 25 top terms are output from each of three components.

For each of these analyses the scikit-learn Python library was used to vectorize the text data, and to perform the relevant dimensionality reduction. For this investigation *variable* represents a lexical item, e.g. a specific noun lemma, or trigrammed noun lemmas. For

each dimensionality reduction method, as in the k-means cluster analysis in the previous section `max_df` was set to 0.05, and `min_df` to 2.

2.7.1 Principal Component Analysis

In Principal Component Analysis (PCA) a multivariate dataset is decomposed into a set of successive orthogonal components that explain the maximum amount of its variance (Pedregosa *et al.*, 2018). PCA is investigated here because it can produce similar components to factor analysis, discussed in the next section, which is a commonly used method in the field of linguistics: for this reason, it was considered of interest to compare the results from these two methods. In addition, PCA is used in conjunction with the word vector similarity method at the end of the current chapter, to reduce word vectors to two dimensions in order to plot them. The PCA process works as follows:

1. find the direction in the data along which the features of that data (variables) are most correlated with each other
2. find the direction in the data that contains the most information while being most orthogonal to the previous component
3. repeat step 2

In this way the least important components can be dropped from the data while the most valuable parts of all the original variables are retained (Müller and Guido, 2016).

The following table lists the top 25 terms from the three components located via PCA. Consideration is also given to what each component is 'about', which is a subjective process. Although only three components were sought in this example, each is relatively meaningful. Components one and two are perhaps about particular arrangements for treatment, while component three, which contains the potentially metaphoric term *journey*, perhaps relates to a more general experience of cancer and the associated

interactions with medical care. The low explained variance shows that three components are insufficient to fully address this data. For example, when the same analysis was run on a sample of 10,000 posts, with the number of components set to 0.1, such that the amount of variance to be explained is greater than 10%, 13 components were returned. While it is typical to set explained variance to 95%, this is less productive with very complex data such as cancer.net, in which every term included is a separate variable.

Table 2-13 principal component analysis output

component	component is about?	top 25 terms from three components	explained variance
1.	visits to medical care	surgery night daughter nurse hubby hour gp problem home end child stage morning scan consultant question couple son case help place tomorrow bed part wife	0.24494854
2.	actual surgery	surgery stage scan surgeon radiotherapy consultant gp biopsy tumour question nurse lymph_node test patient lump oncologist operation experience case symptom type information op option mastectomy	0.16748916
3.	interactions with a gp	gp help feeling situation wife child advice site person thought nurse patient experience symptom heart stage loss word question place information memory grief parent <u>journey</u>	0.11596819

2.7.2 Factor Analysis

Factor analysis uses frequency counts of linguistic features to identify sets of features that co-occur in texts (Biber, 1988, p. 63).

In factor analysis a large number of variables is reduced to a smaller number of factors, each of which combines multiple variables based on the correlations between them.

Factor analysis can produce similar components to PCA, but has the advantage that it is not constrained by the same requirement that all components must be orthogonal to each other (scikit-learn.org, 2021). In linguistics research factor analysis is typically used to

find factors based on the co-occurrence of linguistic variables. In the current investigation factors were sought based on the co-occurrence of linguistic terms.

Although every factor analysis has the same number of factors as it does variables, since the purpose of factor analysis is to reduce a larger number of variables to a smaller number of underlying factors, the aim is to extract as few factors as possible to explain as much of the variation in the data as possible.

The first factor extracts the maximum amount of shared variance, i.e. the largest grouping of co-occurrences in the data; the second factor then extracts the maximum amount of shared variance from the tokens left over after the first factor has been extracted, and so on. In this way, each factor is extracted so that it is uncorrelated with the other factors. ... Factor analysis will continue extracting factors until all of the shared variance among the variables has been accounted for; but only the first few factors are likely to account for a nontrivial amount of shared variance and therefore be worth further consideration. (Biber, 1988, p. 82)

The eigenvalue of a factor is a measure of how much variation in the data that factor explains, with larger values being better, where an eigenvalue is the sum of the squared factor loadings for all the variables in a factor. Because an eigenvalue equal to 1 only accounts for as much variance as a single variable, one method of determining the number of factors is to retain all factors with an eigenvalue >1 (Kaiser, 1960).

If more factors are extracted from the data, more variation in the data is explained. However, factors with smaller eigenvalues are difficult to interpret.

Another method of determining the number of factors to retain is to draw a scree plot of the eigenvalues. All the factors above (to the left of) the inflection point (i.e. the point where the curve starts to level off) are retained, and any factor to the right of the

inflection point is discarded. In reality there can be multiple inflection points, making this a subjective decision.

The current comparison of dimensionality reduction methods is based on use of the same number of components for each method. For this reason, as for the PCA and LDA (subsequent section) investigations, three factors are sought. When the same analysis is run on a sample of 10,000 posts, with number of components set to ten, eight factors have an absolute eigenvalue >1 , such that if the Kaiser method described above was used, eight factors would be sought. As with PCA, although this suggests that three factors are insufficient to more fully explain the cancer.net data, in the terms of the current investigation of this large and very diverse dataset, in which each included term is a variable, the purpose is to find the most dominant factors rather than to explain as much of the variance as possible.

Table 2-14 factor analysis output

factor	factor is about?	top 25 terms from three factors	eigenvalue
1.	eating	couple meal symptom hour nausea med hospital_symptom_tracker control_bag_clothe toothbrush_charger_must load_veggie_bolognese symptom_app_day fatigue_neuropathy_team bed_wall_bedside bed_bag_clothe team_nausea_med toilettry_list_medicine silence_taste chilli_portion_energy cook reminder strain soup table example	0.253
2.	food	calorie cream butter kind powder paste potato food sauce lemon banana coconut pasta teaspoon avocado body skin water egg idea top sugar prawn power challenge	2.678
3.	time and family	night hubby hour daughter morning home weekend couple surgery bed yesterday problem tomorrow nurse child end food christmas son course garden house weather eye body	10.906

As with PCA, the three factor analysis factors listed in the table above can each be interpreted in meaningful ways. Although factor one is a little difficult to interpret as a single coherent topic, it contains references to eating, and to items required for an overnight stay, perhaps in hospital. Factor two seems clearly to be about food, while

factor three has multiple time and social references, and is similar to the k-means cluster two. The factors found in this section are very different to the components found with PCA, and none of them contains any potentially metaphoric term.

2.7.3 Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) is a probabilistic model for finding groups of words that frequently appear together (scikit-learn.org, 2021), which works well for the corpus data used in the current investigation in which each forum post is a separate document. While in the current comparison of LDA with the PCA and factor analysis dimensionality reduction methods only three topics are sought, in the following section, a more in depth use of LDA, the number of topics sought is derived from a processing loop to calculate the coherence score based on different numbers of topics, where the coherence score indicates how likely it is that the terms in a topic do occur together in any particular document. The number of topics used is the number with the highest coherence score.

Table 2-15 LDA output

topic	topic is about?	top 25 terms from three topics
1.	medical issues	surgery nurse gp stage scan consultant radiotherapy question test problem op oncologist symptom biopsy lump side_effect case kind_regard patient information body lung tumour lymph_node experience
2.	feelings and social connections	feeling thought hug child help heart word <u>journey</u> situation place wife person daughter site memory end sister loss world thread home hope son story part
3.	family and time references	hubby night morning tomorrow weekend christmas yesterday daughter hour food hair weather garden couple bed holiday house lol home picture afternoon evening son walk lunch

As with the previous PCA and Factor Analysis investigations, the three LDA topics can be interpreted in a meaningful way. Topic two contains the *journey* token, and is similar to the PCA component containing *journey*, in being a more general and longer-term representation of the experience of living with cancer; for example, it includes the terms *memory*, *hope*, and *story*. The first topic is perhaps about the more specific detail of

medical appointments and is similar to component one found using PCA. Topic three is most similar to the factor analysis factor three, in that it also contains family and time references.

2.7.4 Section summary: using dimensionality reduction to find interpretative repertoires

Dimensionality reduction methods in conjunction with term frequency control were productive in finding meaningful concepts, or interpretative repertoires, in cancer.net.20. Similarities were found between the concepts from each method, and also differences, such that each method was useful in providing more insight into the data. In addition, the potentially metaphoric term *journey* was present in one of the three components for both Principal Component Analysis and Latent Dirichlet Allocation, suggesting again that this could be the primary metaphor in use in this data.

In this section, three different dimensionality-reduction methods were compared in a fairly minimal investigation that sought just three components. In the next section Latent Dirichlet Allocation is used again in a larger-scale more in-depth investigation. Of the three methods investigated above LDA has been selected for the more in-depth investigation because, as discussed above, as a probabilistic model for finding groups of words that frequently appear together (scikit-learn.org, 2021), LDA is well suited to address the very diverse corpus data used in the current investigation in which each forum post is a separate document, and each term is effectively a separate variable.

2.8 Topic modelling with Latent Dirichlet Allocation

Often, when people talk about topic modelling, they refer to one particular decomposition method called Latent Dirichlet Allocation (Müller and Guido, 2016, p. 350)

In topic modelling, terms frequently used together in documents are organised into 'topics', where it is up to the investigator to determine what a topic is about, and if it can be said to be about anything. Murakami et al. (2017) describe topic modelling as a useful bottom-up exploratory approach that may avoid disadvantages of more traditional corpus linguistic techniques, for example the keywords method explored above, which they argue is based on maximum distinctiveness such that it may lead to a form of textual stereotyping. Murakami et al. (2017) also express concerns about topic modelling, however these are addressed in the design of the current study. The variability of topic modelling results based on the definition of texts is not an issue for the current study in which each post naturally forms an individual text for supply to the model. And the issue of the analyst supplying the number of topics to the model, a subjective and possibly distorting input, is addressed by calculating the coherence of the model for different numbers of topics, discussed below.

Topic modelling may provide insights into the concepts and concerns, or in the terms of the current study the interpretative repertoires and ideological dilemmas, of a discourse. For example, in an investigation to predict depression and the strongly associated personality measure of neuroticism from a set of stream of consciousness essays, Latent Dirichlet Allocation (LDA) was found to add value to the prediction of clinical assessments by locating psychologically relevant topics (Resnik, Garron and Resnik, 2013).

Running this more in-depth topic modelling is a way of exploring further:

1. what metaphor themes and potential metaphoremes are present in the data
2. what language is used to express those metaphor themes
3. what metaphoric language typically occurs together

4. what interpretative repertoires are present in the data

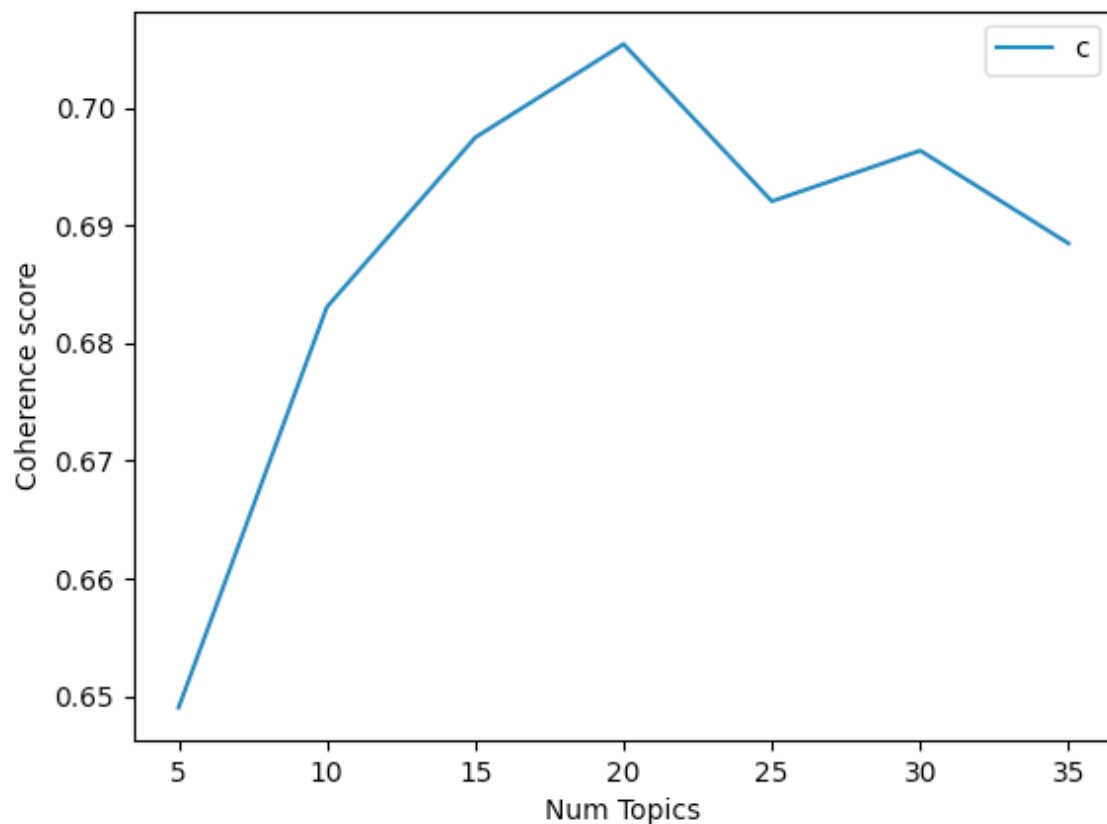
The Gensim Python library *ldamallet* LDA model was used for the investigations in this section.

2.8.1 Determining the optimal number of topics

In topic modelling the coherence score indicates how likely it is that the terms in a topic do occur together in any particular document. The optimal number of topics for a set of data can be estimated, or it can be selected via a processing loop that calculates the coherence of a model with a particular number of topics: the model used is the one with the highest coherence score. For the LDA models below, the number of topics to use was calculated using the Gensim Python library *CoherenceModel* topic coherence pipeline.

The following plot is an example of how the coherence score changes for different numbers of topics; in the example the maximum coherence arises from use of 20 topics.

Figure 2-5 coherence score for different numbers of topics, nouns, trigrams



2.8.2 Visualisation of the LDA output

PyLDAvis (Sievert and Shirley, 2014), an interactive visualisation of the LDA model output, was used to explore the topics found by the model. In the visualisation:

- pale bars represent a term's frequency across the entire corpus
- darker bars represent the frequency of a term in a given topic
- The λ slider is used to adjust the term rankings - small values of λ (near 0) highlight potentially rare, and exclusive, terms for the selected topic; large values of λ (near 1) highlight frequent, but not necessarily exclusive, terms for the selected topic
- the areas of the circles represent that topic's overall prevalence

- the centres of the circles represent the distance between topics (Sievert and Shirley, 2014)

It is suggested, based on an investigative study, that setting λ near 0.6 is helpful for topic interpretation, although this varies across topics and data sets (Sievert and Shirley, 2014). For this reason, λ was set to 0.6 for the results reported in the current study.

2.8.3 LDA topic modelling output

In this section the results are discussed for LDA topic modelling investigations using different parts of speech, and with different max_df settings to focus on terms with different densities in the data.

2.8.3.1 Noun lemmas, trigrams, max_df=0.1, min_df=2

In this analysis only noun lemmas were included, and bigrams and trigrams were formed from them. The *CountVectorizer* was run with a max_df setting of 0.1, such that terms were included that are present in 10% or fewer of posts, and in order to omit idiosyncratic terms the min_df setting was set to 2, such that terms were only included that are present in at least two posts. The results tabled below show that the 15 topics in this LDA output each have a clear focus, which is reflected in the (high) 0.71 coherence score.

Table 2-16 LDA topics with max_df=0.1, min_df=2, λ =0.6, coherence=0.71

topic	topic is about?	terms
1.	family and special occasions	weekend, hubby, christmas, weather, daughter, garden, birthday, son, walk, grandson, morning, work, afternoon, evening, lunch, yesterday, visit, holiday, today, rain, sunday, break, gym, sunshine, saturday, grandchild, present, shop, flower
2.	home concerns	house, tomorrow, night, car, dog, room, tonight, door, bed, tea, tv, box, today, hubby, wall, window, morning, kitchen, clothe, snow, foot, tree, glass, light, cat, pm, card, road
3.	work concerns	work, job, health, nhs, money, form, service, home, benefit, man, carer, idea, charity, company, law, view, country, staff, centre, order, woman, plan, business, office, system, boss, insurance, place, issue, note

4.	specific treatments for cancer	chemo, surgery, radiotherapy, op, side_effect, oncologist, tumour, surgeon, operation, chemotherapy, mastectomy, drug, session, option, cycle, decision, radio, tamoxifen, breast, lumpectomy, reconstruction, grade, hormone, risk, lymph_node, effect, radiation, bra, margin, choice
5.	food	food, weight, water, throat, mouth, luck, recovery, blog, diet, tip, drink, voice, body, progress, tooth, fatigue, tongue, issue, side, stone, tube, amount, neck, problem, medicine, stuff, peg, taste, appetite, tonsil
6.	specific health problems and solutions	bit, hour, eye, night, couple, arm, minute, hand, skin, face, problem, sleep, rest, finger, morning, bag, line, top, min, steroid, thursday, half, back, bed, cream, nail, machine, mark, wound, area
7.	gp interactions	gp, doctor, symptom, blood, scan, infection, lump, ct_scan, anxiety, area, blood_test, biopsy, lymph_node, ultrasound, specialist, sign, cell procedure, test, melanoma, referral, lymphoma, mole, node, case, check, cyst, virus, change, colonoscopy
8.	pain	pain, hospital, nurse, side, hospice, end, medication, call, dr, bone, doc, med, home, chest, problem, tablet, shoulder, ward, pain_relief, morphine, leg, trouble, nerve, stomach, injection, breath, doctor, mo, phone
9.	friends and close relationships	friend, mind, person, partner, feeling, fear, moment, situation, place, illness, future, fact, kind, term, stress, thought, emotion, process, part, conversation, relationship, sense, control, reality, boyfriend, world, reason, space, mam
10.	family and death	dad, child, heart, memory, sister, loss, parent, brother, father, comfort, grief, death, funeral, mum, tear, mom, son, feeling, age, peace, world, grandad, dream, sadness, nan, strength, moment, guilt, father_law, suffering
11.	time references	month, husband, stage, wife, lung, disease, chance, liver, luck, brain, shock, end, march, prognosis, june, bowel, january, october, trial, remission, november, prostate_cancer, august, september, beginning, diagnosis, december, kidney, tumor, quality_life
12.	the forum described more formally as a site of support	support, forum, advice, site, question, experience, information, patient, chat, team, answer, info, situation, condition, group, macmillan, detail, link, website, internet, knowledge, opinion, uk, response, page, diagnosis, type, member, research discussion
13.	family?	today, bit, picture, head, kid, girl, photo, baby, daughter, book, hair, lol, school, tomorrow, holiday, yr, woman, boy, mine, colour, pic, xma, granddaughter, lady, load, cap, wig, colour, rest
14.	concerns around results	result, appointment, consultant, breast, news, kind_regard, diagnosis, test, scan, date, outcome, waiting, wait, mri, biopsy, letter, clinic, mammogram, treatment_plan, regard, lump, worry, breast_clinic, bout_breast_cancer, care_team, finger, touch_kind_regard, consultation, delay, lump_breast
15.	the forum as it operates as a site of support	post, love, hug, <u>journey</u> , hope, thread, word, message, thought, touch, story, news, reply, lady, thinking, strength, max, good, god, prayer, quote, comment, lass, <u>step</u> , inspiration, <u>rollercoaster</u> , <u>road</u> , <u>fight</u> , folk, buddy

Topic 15 contains the potentially metaphoric term *journey*, and the new potentially metaphoric terms *step*, *rollercoaster*, *road*, and *fight*, listed in order of relevance in the topic. The presence of these potentially metaphoric terms together in this topic supports their interpretation as metaphoric, as well as suggesting they may typically be used to discuss the same issues. As discussed in chapter one, the occurrence of different metaphors together suggests that they represent ideological dilemmas: their presence in the same topic here demonstrates that they are typically used to discuss the same issues, regardless of whether multiple metaphoric terms are employed to do this in a single post.

Although both topic 12 and topic 15 appear to be about cancer.net, topic 15 presents a more personal and warm interpretation of the forum, since alongside the more technical terms related to the forum, such as *post*, *thread*, *word*, *message*, *reply*, *quote*, and *comment*, it includes the emotional terms *love*, *hug*, *hope*, *thought*, *strength*, *good*, and *inspiration*, and the familiarly-phrased person references *lady*, *lass*, *folk*, and *buddy*. Use of the potentially metaphoric terms *journey*, *step*, *rollercoaster*, *road*, and *fight*, then, appears to be associated with warmth and familiarity in conjunction with the forum.

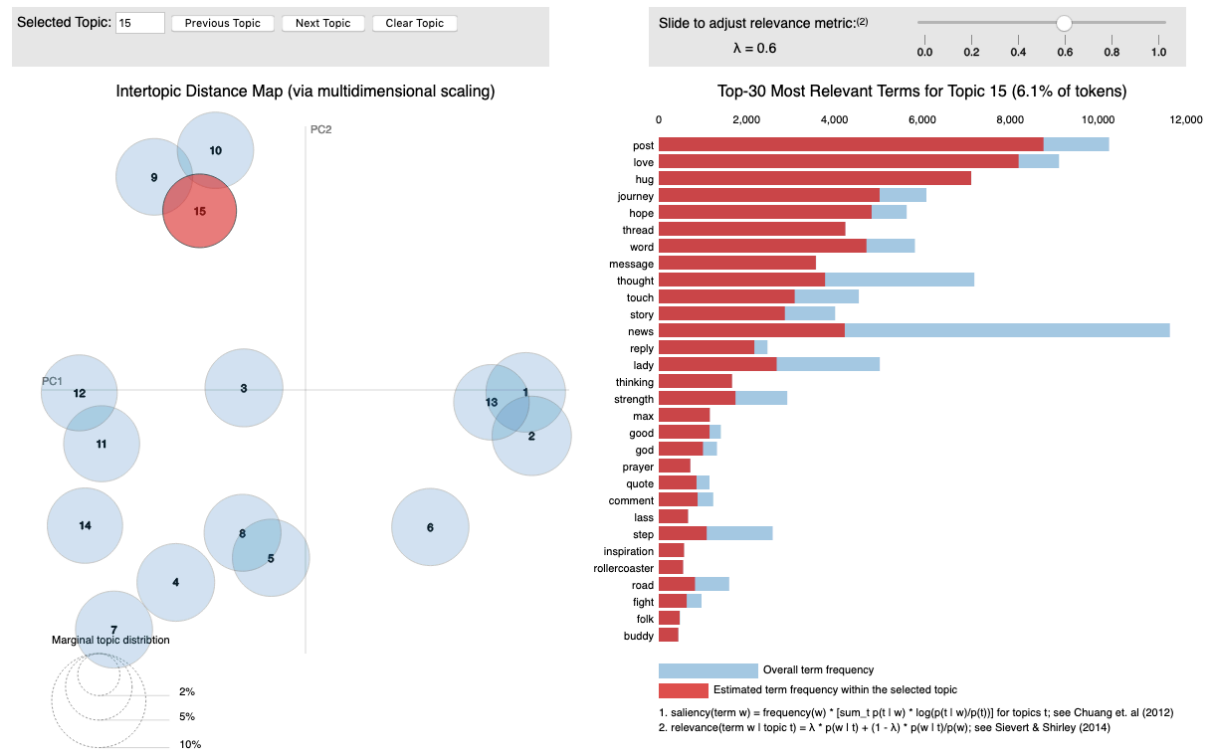
A concordance analysis of the familiarly-phrased person references from topic 15 showed that they have specific uses related to this forum, for example *buddy* is used to address fellow cancer.net participants, *folk* is used in the phrase *virtual folk*, and *lass* is used in the phrase *a fellow breast cancer lass*. This shows that the metaphoric terms in topic 15 support a sense of community on cancer.net.²⁰ beyond its more formal operation as a support community for people living with cancer.

Table 2-17 familiar person references associated with potentially metaphoric terms on cancer.net

term	examples of use
lady	a lovely lady, brave lady, a strong lady, an amazing lady, a wonderful lady
lass	brave lass, an amazing lass, a wonderful lass, a fellow breast cancer lass
folk	caring folk, lovely folk, supportive folk, virtual folk, very kind folk
buddy	my forum buddies, my virtual buddies, be cyber buddies, wonderful virtual buddies, online buddies, neck buddies, breast buddies

A static screenshot of the interactive PyLDAvis plot for topic 15 is shown below.

Figure 2-6 topic 15: a metaphor-related topic



2.8.3.2 Noun lemmas, trigrams, max_df=0.04, min_df=2

To focus more closely on those more rare terms more likely to be metaphoric, for the investigation in this section the max_df setting was set to 0.04, so that terms were included that are present in 4% or fewer posts. The 25 topics in this LDA output each have a clear focus, which is reflected in the 0.70 coherence value. Four of these topics are described in the table below.

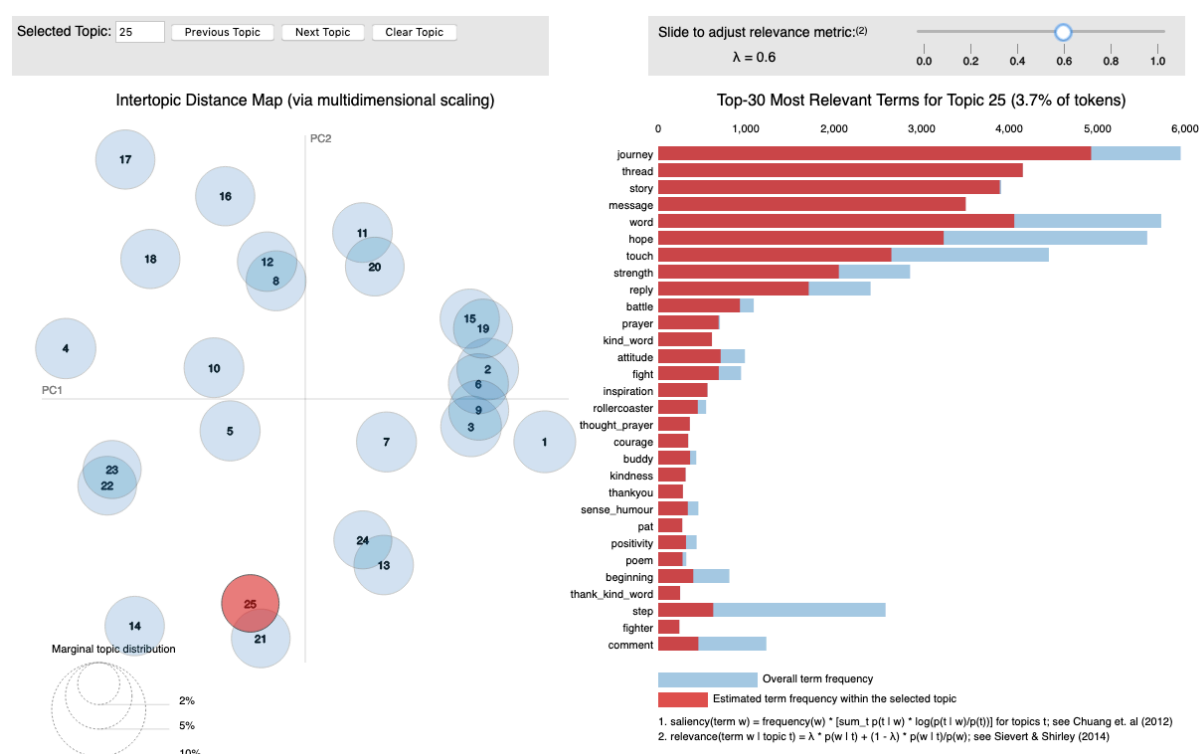
Table 2-18 LDA topics with max_df=0.1, min_df=2, $\lambda=0.6$, coherence=0.70

topic	topic is about?	terms
2	food	food, drink, tea, meal, water, cake, tonight, wine, coffee, alcohol, egg, soup, glass wine, bottle, cheese, sugar, fish, glass, fruit, diet, cup_tea, lunch, chicken, bar
22	anxiety	mind, fear, reason, worry, fact, sort, stress, kind, anxiety, concern, point, answer, part, experience, state, sense, mam, past, health, future, person, panic, idea, control, matter, mistake, chance, positive, distraction, back_mind
23	cancer.net	site, experience, information, chat, info, detail, website, answer, page, link, touch, member, internet, idea, knowledge, type, subject, moderator, contact, discussion, email, regard, situation, response, number, answer_question, reply, list, suggestion
25	the cancer.net community	<u>journey</u> , thread, story, message, word, hope, touch, strength, reply, <u>battle</u> , prayer, kind_word, attitude, <u>fight</u> , inspiration, <u>rollercoaster</u> , thought_prayer, courage, buddy, kindness, thankyou, sense_humour, pat, positivity, poem, beginning, thank_kind_word, <u>step</u> , <u>fighter</u> , comment.

It can be seen from the results table above that topic 25 contains the potentially metaphoric terms *journey*, *battle*, *fight*, *rollercoaster*, *step*, and *fighter*. In comparison to the metaphoric topic 15 in the previous analysis, topic 25 contains the additional potentially metaphoric term *battle* but loses the potentially metaphoric term *road*.

As in the previous analysis, each individual topic makes sense, typically relating to the experience of living with cancer, or relating to use of the forum. But it is the metaphoric topic 25 that evokes the community that is present on cancer.net beyond the factual and descriptive matter of how the forum works. Alongside the terms *story*, *message*, *hope*, *strength*, *prayer*, *kind_word*, *inspiration*, *thought_prayer*, *courage*, *buddy*, *kindness*, *thankyou*, *sense_humour*, *positivity*, and *poem*, the metaphoric terms help to frame the experience of living with cancer as something that is shared by this (caring and supportive) community, and something that is a normal part of a life 'journey', with its inherent change and difficulties.

Figure 2-7 topic 25: a metaphor-related topic



2.8.3.3 Noun lemmas, trigrams, max_df=0.2, min_df=2

To explore further the interpretative repertoires in cancer.net.20, LDA was run again, this time with max_df set to 0.2, so that the focus is less on the more rare potentially metaphoric terms and includes other more common terms from the community interpretative repertoires. The 20 topics in this LDA output each have a clear focus, and this is reflected in the 0.70 coherence value. Five of the topics are described in the table below. Topic 2 contains the terms *peg* and *tube* which are specifically related to the practicalities of eating, as a cancer patient.

Table 2-19 LDA topics with max_df=0.2, min_df=2, λ=0.6, coherence=0.70

topic	topic is about?	terms
1	outside	weather, garden, walk, weekend, car, lunch, sun, shop, afternoon, sunday, morning, house, rain, summer, door, gym, flower, neighbour, sunshine, shopping, dog, bird, tree, window, town, saturday, hubby, colour, snow
2	food, as a cancer patient	food, weight, water, mouth, night, drink, diet, meal, tea, tooth, tongue, cream, stone, peg , taste, appetite, cake, tonight, tube ,

		glass, wine, stuff, bottle, alcohol, egg, dentist, soup, eating, toe, cup_tea
3	practicalities of being a cancer patient	patient, nhs, condition, case, opinion, treatment, decision, team, research, form, service, therapy, prognosis, money, link, benefit, health, information, uk, trial, option, risk, system, charity, cure, disease, centre, country, medicine, issue
	chemo	chemo, treatment, side_effect, hair, oncologist, session, drug, radiotherapy, cycle, week, tablet, radio, effect, dose, chemotherapy, wig, steroid, body, nausea, sickness, round, tiredness, reaction, cap, herceptin, fec, vein, rad, luck, infusion
16	journey	feeling, life, partner, loss, memory, thought, heart, <u>journey</u> , word, emotion, grief, tear, world, comfort, love, part, <u>path</u> , anger, strength, sadness, head, mam, moment, guilt, <u>rollercoaster</u> , soul, dream, shock, thing

Considering more common terms gives a further insight into the *journey* interpretative repertoire (topic 16) as related to emotion, and the *journey* metaphor as providing a way to talk about emotion, since this repertoire includes the emotion-related terms *feeling*, *loss*, *heart*, *emotion*, *grief*, *tear*, *comfort*, *love*, *anger*, *strength*, *sadness*, and *guilt*. The *rollercoaster* potentially metaphoric term is also still present; this is typically used in this context to describe the strongly and suddenly varying emotions that accompany the *journey* of living with cancer. The loss of the *battle* and *fight* terms when the focus is on more common terms shows that these two metaphoric terms are comparatively less dominant in cancer.net.20.

2.8.3.4 Noun, verb, and adjective lemmas, trigrams, max_df=0.2, min_df=2

When LDA was run with the same conditions as the previous section, but including nouns, verbs, and adjectives, the *journey* topic again includes emotion terms (*love*, *hug*, *heart*, *tear*, *feeling*), as well as personal terms relating mostly to children (*kid*, *son*, *baby*, *daughter*, *girl*, *lady*, *child*, *boy*), and the potentially metaphoric terms *journey*, *road*, *fight*, and *path*. It also includes the wide-perspective terms *world*, *god*, and *life*, suggesting that the *journey* interpretative repertoire is used to talk about life as a whole, and perhaps to

bring focus to what is important in life; perhaps to reflect on the writer's own life. The coherence score for this analysis was 0.71.

Table 2-20 LDA topics with max_df=0.2, min_df=2, λ =0.6, nouns, verbs, and adjectives, coherence=0.71

topic	topic is about?	terms
14	journey	love, hug, kid, heart, lot, son, word, hand, baby, <u>journey</u> , daughter, thought, girl, lady, tear, feeling, child, yr, world, <u>road</u> , god, boy, eye, <u>fight</u> , age, smile, head, life, load, <u>path</u>

2.8.4 Section summary: topic modelling with Latent Dirichlet Allocation

Topic modelling in conjunction with term frequency control was used to locate the dominant potentially metaphoric terms in cancer.net. It also showed that metaphoric terms occur together (cluster) in a single interpretative repertoire, suggesting that they are a mechanism for expressing ideological dilemmas, and as such are likely to be a site of important social and psychological work. Topic modelling also provided an insight into the work of the interpretative repertoire of which the potentially metaphoric terms are a part, and therefore the social and psychological work that those terms support, which seems to be related to emotion discourse, and to support a sense of community on cancer.net.

The potentially metaphoric language found so far appears to support two main metaphor themes: *journey*, and *fight*. Although it has also been shown that all of the potentially metaphoric terms tend to be used together, it is anticipated that this represents ideological dilemmas in the relevant interpretative repertoire, such that the metaphoric language discovered still represents separate themes, which are tabled below.

Table 2-21 metaphor themes found in cancer.net.20

metaphor theme	terms used to express the theme
journey	journey, step, rollercoaster
fight	fight, fighter, battle

In the next section word vectors are used to explore language related to these potentially metaphoric terms, to consider what other related metaphoric terms may be used on cancer.net.

2.9 Using word vectors to explore metaphoric language

Word vectors are generated through application of an algorithm to learn word associations from a large corpus of text. Word vector models can be used to find potential near synonyms for words, and to better understand the relationships between words as they are typically used. Each distinct word is represented as a vector of numbers such that the cosine similarity between the vectors indicates the semantic similarity between the words represented by those vectors (Honnibal & Montani, 2021). For this investigation, word vectors from the spaCy en_core_web_lg model were used to explore what terms in the data are most similar to the potentially metaphoric terms already found in the data, to understand the wider meaning of and relationships between those terms, and to look for terms that do similar work. Since the language model is based on common usage, similarly used common metaphoric terms should have semantic similarity even when their literal meaning is not similar.

2.9.1 Vector plot

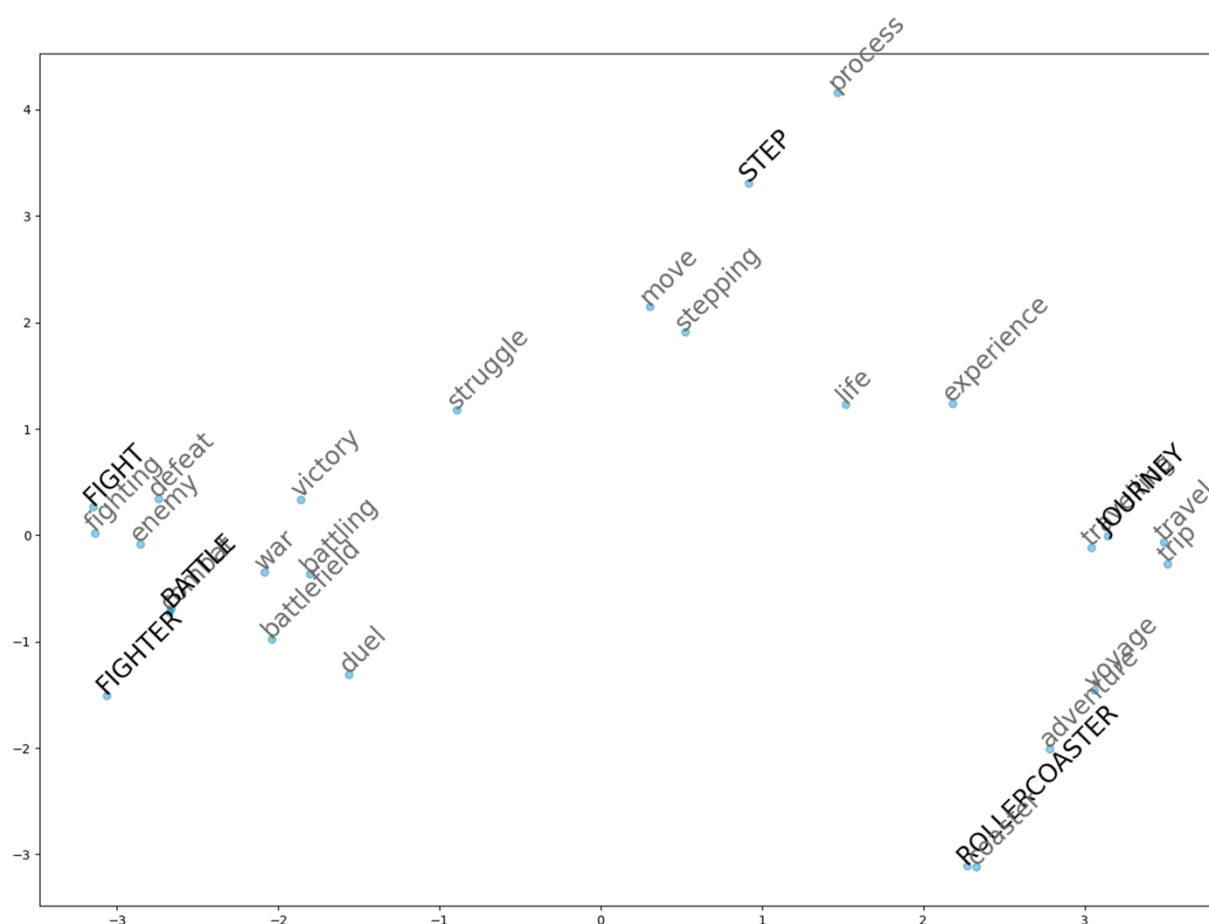
In order to plot the word vectors Principal Component Analysis, discussed above, was used to reduce the vector for each included term to two dimensions.

In the following vector plot tokens already found to have potentially metaphoric significance within the data are written in bold and uppercase so that they can be found

more easily. To constrain the output to focus on those tokens most similar to the potentially metaphoric terms already found, the spaCy word vector similarity method was used such that terms were only plotted that have a similarity > 0.6 to the primary terms under investigation.

Although the plot is still somewhat dense with overlapping terms, most of the terms can be read, and it is helpful to see a visual representation of their semantic proximity. The terms *fight*, *battle*, and *fighter*, are close together at the bottom left of the plot; the terms *journey* and *rollercoaster* are positioned close together at the bottom right of the plot; the term *step* in comparison, at the centre top of the plot, is not close to any of the other potentially metaphoric terms. This provides information about the work of different terms in invoking metaphor, and which terms perform a similar function in common usage. This separation supports the view in the previous section that different metaphor themes in the same interpretative repertoire represent ideological dilemmas.

Figure 2-8 vector plot of nouns found to have potential metaphoric significance plus other similar nouns



To make the data easier to read, the following table shows all nouns from the plot in descending order of similarity for each term. Where a new term has a similarity > 0.6 with more than one of the original potentially metaphoric terms it has been grouped with the term with which it has the highest similarity. Each of the terms in this table is explored in more detail in chapter three, to ascertain its actual metaphoricity in cancer.net.

Table 2-22 potentially metaphoric terms and their similar nouns

potentially metaphoric term	similar noun	similarity (>0.60)
battle	defeat	0.69
battle	war	0.65
battle	combat	0.64
battle	battlefield	0.63
battle	duel	0.62
battle	enemy	0.61

battle	struggle	0.61
battle	victory	0.60
journey	travel	0.66
journey	trip	0.65
journey	voyage	0.65
journey	adventure	0.64
journey	life	0.63
journey	travelling	0.62
journey	experience	0.60
rollercoaster	coaster	0.81
step	process	0.66
step	stepping	0.63
step	move	0.61

2.9.2 Using vector norms to find dissimilar concepts in a text

A vector norm is the L2 norm of a token's vector (the square root of the sum of all values, squared), which represents the vector as a single value (Honnibal and Montani, 2021). It is fundamental to metaphoric language that dissimilar concepts are used together, with a source concept typically being used to frame an understanding of a target concept. Vector norms can be used to highlight when dissimilar concepts are used together, for example by calculating the distance between the vector norms of two tokens or calculating the standard deviation of vector norms in a text.

In the following investigation of all the posts from one cancer.net.20 participant, the standard deviation of the vector norms for all nouns in each post was calculated. The following post containing metaphoric language was found to have the highest standard deviation ($\sigma=1.455$), suggesting that it contains dissimilar nouns. To show which nouns contributed to this high standard deviation those tokens that were included in the calculation are listed below the post text, along with their vector norm. It should be noted

that proportionally dominant use of the word *cancer* in this post in conjunction with the other nouns increases the standard deviation, while terms already identified as potentially metaphoric, which are only present in this post in verb form, are not actually included in the calculation.

Table 2-23 post with highest standard deviation of vector norms of all nouns

hay. that's amazing. you did good here's to kicking cancers butt all the way you rock.

The following tokens from this post are included in the calculation of the standard deviation of vector norms:

Table 2-24 vector norms for nouns included in calculation of standard deviation of vector norms

token	vector norm
way	4.7604
good	5.2805
cancers	8.0736

Of all posts from the same participant that contain the terms already identified as potentially metaphoric, the vector norms of the nouns in the following post were found to have the lowest standard deviation ($\sigma=0.489$). This fairly long post has been truncated for privacy reasons, but all nouns included in the calculation are listed below the post.

Table 2-25 post with lowest standard deviation of vector norms of all nouns

i know what you mean about the feeling of 'not real' ... it's a bumpy ride but if you look around we're all on that same ride with you.

The following tokens were included in the calculation of the standard deviation of vector norms for this post:

Table 2-26 vector norms for nouns included in the calculation of standard deviation of vector norms

token	vector norm	token	vector norm
given	4.9030	reach	5.5850
mean	4.9843	try	5.6073
others	4.9988	problem	5.7470
ones	5.0839	inspiration	5.8137
time	5.1279	downs	5.8171
give	5.1312	bit	5.8507
know	5.1606	help	5.8787
ups	5.1821	share	5.9220
hold	5.2284	days	5.9896
talking	5.2475	jumping	6.0184
let	5.2976	advice	6.0325
look	5.3037	hug	6.3495
keep	5.3458	feeling	6.3642
holding	5.3478	<u>journey</u>	<u>6.3805</u>
talk	5.4693	<u>rollercoaster</u>	<u>6.3836</u>
lots	5.4869	ride	6.5781
day	5.5376	grade	6.6439

Although the potential metaphor source terms *journey* and *rollercoaster* are present in this text, their target is not made explicit: since the *cancer* token is not present, the dissimilarity that supports metaphoric interpretation is implicit rather than explicit, and this explains the relatively low standard deviation for this metaphoric post. In addition, although the literal terms *journey* and *rollercoaster* are not very similar, their similar metaphorical use is represented in their very similar vector norms, which are underlined in the table above. This exemplifies an issue with a selectional preference violation approach to finding metaphor in text: "in the case of frequent conventional metaphors, no statistically significant violation can be detected in the data" (Veale, Shutova and Klebanov, 2016, p. 90). At the same time this issue supports the use of the similarity method in the previous section for finding additional metaphoric language relating to community metaphor themes.

2.10 Chapter discussion

In this chapter corpus linguistic and data science techniques were used to prospect for metaphors, and their associated interpretative repertoires, to provide an insight into the work of metaphors in this community.

When using topic modelling to prospect for metaphors the focus was initially on less frequently used nouns that are present in 10% and 4% or fewer of documents. In this case potential metaphors with a different source domain were located in the same interpretative repertoire, suggesting that those metaphors represent ideological dilemmas, and that this applies even where the metaphoric terms do not occur together in the same post: their presence in the same interpretative repertoire demonstrates that they are typically used to discuss the same issues through a different framing. In addition, the high coherence score obtained with every setting of `max_df` investigated supports the view of cancer.net as a community, since it demonstrates that the concerns of the forum are clear and coherent.

When more prevalent nouns were included in the topic output that are present in 20% or fewer of documents, the interpretative repertoire in which *journey* and *rollercoaster* are located appears to support discussion of emotion. When more common nouns, verbs, and adjectives were included, use of the *journey* metaphor appears to support use of emotion terms, personal terms, and the wider-focus terms *world*, *god*, and *life*, suggesting that the interpretative repertoire containing *journey* is used to talk about life as a whole, and about concepts that are wider than a single life.

Word vectors were then used to identify other nouns in the data that are used similarly to nouns already established as potentially metaphoric: to find additional potentially metaphoric nouns.

In the next chapter qualitative investigations are undertaken to evaluate the metaphoricity of each of the potentially metaphoric terms located in the current chapter, which are summarised in the table below. Investigations are also made into signalled metaphor in cancer.net, and its relationship to the established metaphor themes of this community.

Table 2-27 potentially metaphoric nouns and the nouns most similar to them in cancer.net.20

potentially metaphoric noun	similarly used nouns from cancer.net.20
journey	travel, trip, voyage, adventure, life, travelling, experience
step	process, stepping, move
rollercoaster	coaster
fight	fighter
battle	defeat, war, combat, battlefield, duel, enemy, struggle, victory

3 EVALUATING POTENTIAL METAPHOR

3.1 Introduction

In this chapter qualitative and quantitative methods are combined to build on the insights gained in chapter two, in which potentially metaphoric nouns from cancer.net.20 (all posts from participants who have made 20 or more posts) were gathered without preconception of what they might be, i.e. without searching for specific terms, through the use of corpus linguistic and unsupervised data science methods. The potentially metaphoric nouns collected in chapter two are listed in the following table alongside similarly used nouns from cancer.net.20 collected via the word vector similarity method.

Table 3-1 potentially metaphoric nouns and nouns most similar to them in cancer.net.20

potentially metaphoric noun	similarly used nouns from cancer.net.20
journey	travel, trip, voyage, adventure, life, travelling, experience
step	process, stepping, move
rollercoaster	coaster
fight	fighter
battle	defeat, war, combat, battlefield, duel, enemy, struggle, victory

In this chapter these potentially metaphoric terms are considered qualitatively within their local context to determine their actual metaphoricity, and their identity as characteristic metaphors, and metaphoremes, of the cancer.net community. In addition, typical linguistic constructions used to signal metaphor are explored, to gain more insights into the use of metaphor in conjunction with the experience of living with cancer, and to consider the relationship between signalled metaphor and characteristic community metaphor. At the end of this chapter evaluated cancer.net metaphor themes are summarised to support an analysis of language style in relation to metaphor use in chapter four.

corpus linguistics makes it possible to combine qualitative and quantitative analysis in an explicit and rigorous manner, and is therefore ideally suited to bridge the methodological divide in healthcare research (Semino et al., 2017, p. 13).

The first step in this chapter is a discussion of metaphor identification procedures, following which a rationale is made for the method to evaluate metaphor used in the current study. After this, concordance analysis is used to consider the potentially metaphoric terms gathered in chapter two in their context, and to calculate the proportion of metaphoric and non-metaphoric uses of each term. Next, an exploration is undertaken of metaphoric noun phrases, and their identity as metaphoremes is considered. To support this analysis, graph-based ranking is used to rank metaphoric noun phrases based on their connectedness to other relevant terms within the text. Subsequently, a comparison of the density of metaphoric phrases is made between cancer.net.20 (all posts of participants who have made 20 or more posts), and cancer.net.LT20 (all posts of participants who have made fewer than 20 posts). Finally, typical linguistic constructions used to signal metaphor are explored, and a quantitative analysis of the relationship between dominant metaphor themes of the cancer.net community and the various types of signalled metaphor provides further insight into the social and psychological work of metaphor in this context.

The outcome for this chapter is the establishment of a set of evaluated dominant metaphoric terms for cancer.net, organised into themes, which are explored further in terms of associated language style in chapter four, and in a diachronic analysis in chapter five.

3.2 Metaphor identification procedures

In this section formalised metaphor identification procedures are considered, and a rationale is made for the identification method used in the current study.

3.2.1 MIP

The MIP (Metaphor Identification Procedure), a standardised process to reliably find linguistic metaphor, was established to make metaphor research more empirically sound, and more comparable (Pragglejaz, 2007). The Macmillan online dictionary is used to check word meanings because it is a systematically processed corpus which is large enough to provide multiple citations for all but the most rare words. It is also relatively recent, and aims to provide a description of current English (Pragglejaz, 2007), which is most suitable for the cancer.net web corpus used in the current investigation.

The default position of the MIP is to treat each word as a separate unit, such that multiword units are considered together only when they can only be understood as a whole. Word class may be ignored in MIP, which consists of the following steps:

Figure 3-1 MIP Metaphor Identification Procedure steps

1. Read the entire text–discourse to establish a general understanding of the meaning.
2. Determine the lexical units in the text–discourse.
3. For each lexical unit
 - a. establish its meaning in context, that is, how it applies to an entity, relation, or attribute in the situation evoked by the text (contextual meaning). Take into account what comes before and after the lexical unit.
 - b. determine if it has a more basic contemporary meaning in other contexts than the one in the given context. For our purposes, basic meanings tend to be

- i. More concrete; what they evoke is easier to imagine, see, hear, feel, smell, and taste.
 - ii. Related to bodily action.
 - iii. More precise (as opposed to vague)
 - iv. Historically older.
4. Basic meanings are not necessarily the most frequent meanings of the lexical unit.
 5. If the lexical unit has a more basic current–contemporary meaning in other contexts than the given context, decide whether the contextual meaning contrasts with the basic meaning but can be understood in comparison with it.
 6. If yes, mark the lexical unit as metaphorical. (Pragglejaz, 2007)

3.2.2 MIPVU

Taking MIP as its starting point, MIPVU was developed to further refine the choices MIP entails, with VU added as a reference to the university where the technique was defined: Vrije Universiteit Amsterdam (Steen *et al.*, 2010). In contrast to MIP, in MIPVU it is not permissible to cross word class boundaries when considering the basic meaning of a lexical unit. For example, the contextual meaning of a verb cannot be compared to its basic meaning as a noun. In addition, direct metaphor, indirect metaphor, borderline cases of metaphor, metaphor signals, and metaphor due to personification, described in more detail in the figure below, are all included in MIPVU but not MIP.

Figure 3-2 metaphor types included in MIPVU but not in MIP

1. Direct metaphor, in which the contextual meaning is the same as the more basic meaning, with a comparison being expressed through direct language use that may or may not be signalled. For example: he’s like a ferret.

2. Indirect metaphor, in which there is an underlying grammatical or semantic link that creates the metaphorical interpretation.
3. Borderline cases of metaphor: the WIDLII tag (When In Doubt, Leave It In) is used to recognise borderline cases whose status can't be resolved by the analyst independently, or through discussion with the research group.
4. Metaphor signals, in which metaphor can be signalled by metaphor markers such as like, as, as if, so-called, resembling etc. More general signals such as sort of or kind of are not included in MIPVU.
5. Metaphor due to personification, in which metaphorical tension is based on a comparison between the human and the non-human. (Steen *et al.*, 2010)

3.2.3 DMIP

A metaphor is potentially deliberate when the source domain of the metaphor is part of the referential meaning of the utterance in which it is used (Reijnierse *et al.*, 2018, p. 134)

DMIP (Deliberate Metaphor Identification Procedure) is a method for the systematic and reliable identification of potentially deliberate metaphor in language use, in which the source domain provides "an alien or alternative perspective" (Reijnierse *et al.*, 2018, p. 133) for the target domain. The distinction between metaphors that are deliberately used as metaphor, and those that are not, is described as relating to communicative intention.

The central feature of deliberate metaphor is said to be attention to the source domain as a distinct domain of reference. For example, in the phrase *The political battlefield is strewn with corpses*, the metaphoric terms *battlefield* and *corpses* are identified as potentially deliberate because they each introduce a new perspective on the target domain POLITICS from the source domain WAR. The phrase *The Battle for New York's Key Voting Blocs in the Primaries* similarly describes the target domain POLITICS in terms of the source

domain WAR. However according to DMIP, *battle* in this context is not identified as potentially deliberate, because it does not explicitly introduce a different perspective on the target domain POLITICS in the meaning of the phrase (Reijnierse *et al.*, 2018).

DMIP, since it focuses specifically on deliberate metaphor, is not a general metaphor identification procedure. After MIPVU is applied to find all metaphor related words (MRW), in DMIP the extra step described in the figure below is used to consider whether an MRW is potentially deliberate.

Figure 3-3 DMIP additional steps after MIPVU

Determine whether the source domain of the MRW is part of the referential meaning of the utterance in which the MRW is used.

- a. If yes, mark the MRW as potentially deliberate, and note how this is the case
- b. If no, mark the MRW as non-deliberate
- c. In case of doubt, mark the MRW as potentially deliberate, and add the WIDLII tag
(when in doubt leave it in)

3.2.4 Discourse Dynamics Framework

In comparison to the identification methods discussed above, in the discourse dynamics framework the focus of interpretation is not at the word level, but on stretches of language that might be metaphorical, since when language is used to express thoughts and ideas these are understood as being adjusted as they are used, for more effective communication of meaning, for example in the concept of metaphor shifting discussed in chapter one. In the discourse dynamics framework, in order to determine whether a word or phrase is metaphoric it is considered against the following two requirements:

Figure 3-4 discourse dynamics framework process for determining the metaphoricity of a word or phrase

1. there is a contrast or incongruity between the meaning of the word or phrase in its discourse context and another meaning, together with
2. there is a transfer of meaning that enables that contextual meaning to be understood in terms of basic meaning (Cameron and Maslen, 2010)

3.2.5 Metaphor identification method used in the current investigation

In the current investigation MIPVU is used to identify metaphor, since it refines and clarifies the choices entailed in use of the MIP, and because it includes signalled metaphor, which is a particular focus later in this chapter. MIPVU also does not allow the crossing of word class boundaries when considering the basic meaning of a lexical unit, and this aligns with the method of the current investigation in which the focus is on metaphors instantiated as nouns.

The focus of the current study is on potentially more deliberate metaphor which, as discussed previously, is understood as being more adaptable to specific community purposes over time in comparison to more instinctive and less visible metaphor such as the conduit metaphor. The focus on potentially more deliberate or active metaphor in the current investigation is, as discussed in chapter two, based on the consideration of dominant metaphors instantiated as nouns, which it has been argued are the most active and accessible of metaphors. Although it is not possible to say if a linguistic metaphor is or is not deliberate, the analysis considering use of signalled metaphor in conjunction with community metaphor themes later in this chapter provides insights into the potentially deliberate use of metaphor in this context.

The mixing of metaphor is also related to deliberateness in the current investigation, where as discussed previously in relation to ideological dilemmas (section 1.5.3.1) the

mixing of metaphor has been argued to be an instantiation of potentially more deliberate metaphor (Steen, 2016). The perception that metaphors are being mixed, it is argued, depends on the interpretation or experience of those metaphors as deliberate: "all other cases of conceptual clashes between adjacent metaphors do not get recognized as mixed metaphor because they are not used deliberately as metaphors" (Steen, 2016, p. 114). From a dynamic perspective this may be related to activation levels of metaphor, where "the mixing of metaphors changes the semantic salience structure, creating different version and degrees of activated metaphoric meaning" (Müller, 2016, p. 32). In comparison to the discourse dynamics framework discussed earlier (Cameron et al., 2009), which has a focus on potential metaphoricity of linguistic metaphors for participants in a conversation, Müller (2016) relates the mixing of metaphor to display by participants of the presence of metaphoric meaning in their communicative intent. In terms of the career of metaphor theory (Bowdle & Gentner, 2005), the mixing of metaphor perhaps reinstates cross-domain mapping (comparison) when in metaphor in comparison to simile there is typically within-domain mapping (categorisation), which may be more conventional, and therefore potentially less deliberate. In terms of health discourse specifically, Charteris-Black (2016) similarly finds that the mixing of metaphor supports the communication of pain, specifically occurring where the intent is to emphasise the intensity of pain, and to characterize it as out of control. Where aspects of pain are referred to as controllable, in comparison, metaphor use is more semantically convergent, such as repeated and extended metaphors based on the same source domain. The mixing of metaphor may also be related to blending theory, discussed in section 1.3.2, in which multiple conceptual structures are combined, which may be essential to the explanation of how complex mixed and extended metaphor may be understood.

This leaves the related issue of whether a metaphor is received and interpreted as metaphoric by any particular reader at any particular reading, into which the consideration of metaphor use in conjunction with language style in subsequent chapters may provide an insight.

Recognition of the dynamic nature of metaphor use and interpretation described in the discourse dynamics framework is also an important consideration in investigating the work of metaphor in cancer.net. While the framework can not be applied in a meaningful way to the context of the individual sentences and phrases that are under consideration in the current chapter, the dynamism of the cancer.net metaphor themes is a focus in the diachronic analysis in chapter five of the current study.

In the next section the metaphoricity of the potentially metaphoric terms collected in chapter two is assessed, and their use in cancer.net.20 is quantified.

3.3 Investigation into metaphoricity of terms

For each potentially metaphoric term collected in chapter two, to determine the proportion in which that term is used metaphorically in this context 20 instances of the lemma of each term as a noun were sampled as a Sketch Engine concordance. The results tables summarise for each potentially metaphoric term as a noun: the number of instances of that term; its density per one million tokens; and the proportion of analysed occurrences that are metaphoric in their context. Metaphoric examples are also provided to build a sense of the use, and associated work, of each metaphoric term in this context, as well as examples of literal use where relevant, which are crossed out.

In addition, for the primary potentially metaphoric terms found via the topic modelling investigation in chapter two a comparison is made between data for participants who have

made 20 or more posts (cancer.net.20), and those who have made fewer than 20 posts (cancer.net.LT20). This gives further insight into the extent to which a particular metaphoric term is characteristic of the discourse of this community, since such a metaphor is expected to be used more frequently by more established participants than by less established participants. All other terms are investigated solely on the data for participants who have made 20 or more posts (cancer.net.20).

3.3.1 Dictionary definition of terms

For each of the potentially metaphoric terms located through topic modelling in chapter two, listed in the table below, the first Macmillan definition of that term as a noun is basic in the sense of being concrete; related to bodily action; and precise in terms of context (Pragglejaz, 2007).

Table 1 potentially metaphoric nouns from cancer.net.20 found via topic modelling

journey	step	rollercoaster	fight	fighter	battle
---------	------	---------------	-------	---------	--------

The Macmillan dictionary also lists a metaphoric definition for each of these terms, demonstrating that they are commonly used with a metaphoric sense. However, each of those metaphoric senses (underlined in the tables below) still has an active metaphorical basis that is available to frame the target concept by accentuating some of its features and suppressing others. In contrast the metaphorical mapping of the term *pedigree*, which is derived from the French term for the foot of a crane (pied de grue) for example "is no longer current to contemporary speakers and thus pedigree is a true 'dead metaphor'"(Pragglejaz, 2007, p. 30): when the term *pedigree* is used currently, it does not commonly make available the concept of a crane's foot, except perhaps to researchers of metaphor.

In comparison to the potentially metaphoric terms found through topic modelling, some of the related nouns found through the word vector similarity method in chapter two do not have a basic definition, but only more abstract and much more general definitions. For example, *adventure*, located because of its similar meaning to *journey* in general usage, is defined in the Macmillan dictionary as "an exciting, unusual, and sometimes dangerous experience". No metaphoric instances were found of such terms in this context.

In the following sections, the Macmillan definitions for each potentially metaphoric term as a noun are used to determine whether or not a particular situated use of a term is metaphorical or not.

3.3.2 journey and related terms

The table of Macmillan definitions below shows that *journey* and its related terms *travel*, *trip*, *voyage*, and *life*, all have a basic first meaning. Subsequent metaphoric definitions of these terms are underlined. In comparison the terms *adventure*, and *experience*, do not have a basic meaning.

Table 3-2 Macmillan definitions for journey and related terms; metaphoric definitions are underlined

term	definition as a noun	has basic meaning?
journey	<ul style="list-style-type: none"> an occasion when you travel from one place to another, especially when there is a long distance between the places <u>a process of changing and developing over a period of time</u> 	yes
travel	<ul style="list-style-type: none"> the activity of travelling (travel=go on a long journey) 	yes
trip	<ul style="list-style-type: none"> an occasion when you go somewhere and come back again 	yes

	<ul style="list-style-type: none"> • an occasion when you fall because you hit your foot on something • <u>a strange experience that someone has because they have taken a powerful illegal drug</u> • <u>an occasion when you experience a strong emotion</u> • <u>an unusual and enjoyable experience</u> 	
voyage	<ul style="list-style-type: none"> • a long journey, especially by boat or into space 	yes
adventure	<ul style="list-style-type: none"> • an exciting, unusual, and sometimes dangerous experience 	no
life	<ul style="list-style-type: none"> • time from birth to death • way of living, experience • state of being alive • time something exists/lasts • <u>activity/excitement</u> • life imprisonment 	yes
experience	<ul style="list-style-type: none"> • knowledge and skill that is gained through time spent doing a job or activity • the knowledge that you get from life and from being in a lot of different situations • something that happens to you, or a situation you are involved in 	no

The following table summarises for *journey* and its related terms: the number of instances there are of that term as a noun; its density per one million tokens; and the proportion of analysed occurrences that are metaphoric in their context. Metaphoric examples are also provided to build a sense of the use, and associated work, of each metaphoric term in this context, as well as examples of literal use where relevant, which are crossed out. In addition, for *journey* a comparison is made between data for participants who have made 20 or more posts (cancer.net.20), and those who have made fewer than 20 posts (cancer.net.LT20). All other terms are investigated solely on the cancer.net.20 data.

It can be seen from the results table below that *journey* is used predominantly metaphorically in this context for both participant groups, but the other, related, terms are not used predominantly metaphorically. An example of metaphor mixing, which as discussed above may be deliberate use of metaphor, is marked with an asterisk, and reference to *journey* in conjunction with *cancer* as the phrase *cancer journey* meets the DMIP criterion for potential deliberateness described above. Non-metaphoric examples are crossed out.

Table 3-3 journey and related terms analysis of metaphoricity

term	count	density	metaphoric instances	metaphoric examples
journey cancer.net.20	5,580	332.23	19/20 (96%)	*it's not about winning or losing this journey of ours my own cancer journey a plane journey away
journey cancer.net.LT20	2,603	162.73	18/20 (90%)	happiness is a journey, not a destination
travel		24.77	0/20	
	416			
trip	1,391	82.82	0/20	
voyage	13	0.77	2/13 (15%)	my hopeful voyage to find some kind of sanity my voyage to an unknown place bon voyage
adventure	74	4.41	0/20	
life	18,705	1,113.69	0/20	
experience	4,694	279.48	2/20 (10%)	the voice of experience their up and down experiences

3.3.3 step and related terms

The table of Macmillan definitions below shows that *step* has a basic first meaning, and subsequent metaphoric definitions, which are underlined. The related term *process* does not have a basic first meaning, and while the second meaning for *process* is basic in that it refers to a physical object, this more obscure technical definition is not accepted for the

purposes of the current investigation. The first meaning of *move* is abstract, however subsequent meanings are more basic.

Table 3-4 Macmillan definitions for step and related terms; metaphoric definitions are underlined

term	definition as a noun	has basic meaning?
step	<ul style="list-style-type: none"> • a short movement made by putting one foot in front of the other • a particular movement or set of movements performed with your feet when you are dancing • the particular way that someone walks, that sometimes shows how they are feeling • <u>one of a series of actions that you do in order to achieve a particular aim</u> 	yes
process	<ul style="list-style-type: none"> • a series of things that have happened and have a particular result • an official document ordering you to take part in a case of court law (technical) 	no
move	<ul style="list-style-type: none"> • <u>something that you do in order to achieve an aim or solve a problem, often one of a series of things</u> • a change in the place where you live or work • a change in an activity, career, situation • a change in the position of one of the objects used in board games 	yes

It can be seen from the results table below that *step* and *move* are predominantly used metaphorically in this context, but *process* is not used metaphorically. An example of *step* used in conjunction with *journey* is marked with an asterisk.

Table 3-5 step and related terms analysis of metaphoricity

term	count	density	metaphoric instances	metaphoric examples
step	1,989	118.42	17/20 (85%)	i'm walking in your footsteps, only half a step behind i can also see your step dad's side
cancer.net.20				
step	1,201	94.74	16/20 (80%)	*the only thing that is ultimately real about your journey is the step you are taking at this moment i fell down one step
cancer.net.LT20				
process	1,980	117.89	0/20	
move			12/20 (60%)	you must acknowledge they have gone and move on move to the country and become a chicken farmer

3.3.4 rollercoaster and related terms

The table of Macmillan definitions below shows that *rollercoaster* and its related term *coaster* both have a basic first meaning. In addition, *rollercoaster* has a metaphoric definition, which is underlined.

Table 3-6 Macmillan definitions for rollercoaster and related terms

term	definition as a noun	has basic meaning?
rollercoaster	<ul style="list-style-type: none"> a structure like a tall railway with steep slopes at an amusement park, that you have fast rides on for fun <u>a situation in which there are many big and sudden changes</u> 	yes
coaster	<ul style="list-style-type: none"> a small flat object that you put under a cup to protect the table a ship that travels along the coast from one port to another carrying goods a roller coaster (American) 	yes

The results table below confirms that *rollercoaster* is always used metaphorically in this context, for both more established and less established participant groups, but *coaster* is occasionally used literally. However, *coaster* is always used metaphorically when used in

conjunction with *roller*. An example of *rollercoaster* used in conjunction with *journey* is marked with an asterisk.

Table 3-7 rollercoaster and related terms analysis of metaphoricity

term	count	density	metaphoric instances	metaphoric examples
rollercoaster cancer.net.20	822	48.94	20/20	we're right on the rollercoaster with you it's a rollercoaster of emotions
rollercoaster cancer.net.LT20	362	28.56	20/20	it really is an emotional rollercoaster
coaster	284	16.91	19/20 (95%)	*many of us on here describe the cancer journey as a roller coaster an old coaster and tea light holder

3.3.5 fight and related terms

The Macmillan definitions in the table below show that *fight* and its related term *fighter* both have a basic first meaning; they also each have at least one metaphoric definition, underlined in the table.

Table 3-8 Macmillan definitions for fight and related terms

term	definition as a noun	has basic meaning?
fight	<ul style="list-style-type: none"> • hitting and kicking • between soldiers • <u>disagreement/argument</u> • <u>competition to win something</u> • <u>attempt to stop/achieve</u> • <u>determination to try</u> 	yes
fighter	<ul style="list-style-type: none"> • a military plane that is designed for battles with other planes • someone who takes part in the sport of boxing or any other sport in which people fight physically • someone who often gets involved in physical fights • <u>someone who refuses to be defeated even in the most difficult situations</u> 	yes

It can be seen from the results table below that *fight* and *fighter* are predominantly used metaphorically in this context. It is notable that the density of *fight* is 61.45 for cancer.net.20, and 59.79 for cancer.net.LT20: there is no notable difference in *fight* density between these different groups of participants. This is discussed further at the end of this section.

Table 3-9 fight and related terms analysis of metaphoricity

term	count	density	metaphoric instances	metaphoric examples
fight cancer.net.20	1,032	61.45	19/20 (95%)	it still hits you like a cannonball when the fight is lost have just broken up a fight outside my back door between 7 or 8 magpies
fight cancer.net.LT20	758	59.79	19/20 (95%)	i very happily accept our pink boxing gloves because i am ready for a huge fight!!!
fighter	207	12.32	20/20	he's a fighter and he will be fine I know it! people used to say i was strong or a fighter

3.3.6 battle and related terms

The Macmillan definitions in the table below show that *battle* and all its related terms except *defeat* have a basic first meaning; each term but *defeat* also has at least one metaphoric definition, underlined in the table.

Table 3-10 dictionary definition for battle and related terms

term	definition as a noun	has basic meaning?
battle	<ul style="list-style-type: none"> a fight between two armies in a war <u>a situation in which different people or groups compete with each other in order to achieve something or get an advantage</u> <u>a situation in which someone is trying very hard to deal with a difficult situation</u> 	yes
defeat	<ul style="list-style-type: none"> failure to win a competition or to succeed in doing something 	no
war	<ul style="list-style-type: none"> when countries fight <u>effort to stop something</u> when people compete 	yes

	<ul style="list-style-type: none"> • when people fight/argue • Second World War 	
combat	<ul style="list-style-type: none"> • fighting during a war • <u>an attempt to stop something bad or to solve a difficult problem</u> 	yes
battlefield	<ul style="list-style-type: none"> • a place where a battle takes place or where one took place in the past • <u>a situation in which people disagree and cause problems for each other</u> 	yes
enemy	<ul style="list-style-type: none"> • someone who is opposed to someone else and tries to do them harm • <u>something that harms or threatens someone or something</u> 	yes
struggle	<ul style="list-style-type: none"> • <u>an attempt to do something that takes a lot of effort over a period of time</u> • a fight or a war • something that takes a lot of physical effort 	yes
victory	<ul style="list-style-type: none"> • the fact of winning a competition or battle, or an occasion when someone wins • <u>a situation in which someone's principles or goals become officially accepted</u> 	yes

The concordance analysis results table below shows that *battle*, *battlefield*, and *struggle*, are predominantly used metaphorically in this context, but *defeat*, *war*, *combat*, *duel*, *enemy*, *struggle*, and *victory*, are not. An example of metaphor mixing is marked with an asterisk.

Table 3-11 battle and related terms analysis of metaphoricity in cancer.net

term	count	density	metaphoric instances	metaphoric examples
battle cancer.net.20	980	58.35	20/20	*when my husband was diagnosed he chose journey because he felt battle was inappropriate as there was no chance to 'win' it in his circumstances the mental battle is a very tough one
battle cancer.net.LT20	987	77.86	20/20	a very short battle with pancreatic cancer
defeat	23	1.37	0/20	the consultant admitted defeat and said that the left implant was rejecting again
war	202	12.03	4/20 (20%)	cancer has not won the war the first gulf war a world war 1 medal
combat	3	0.18	2/3 (66%)	help combat side effects take medication to combat chemo problems going to try body combat and get rid of my frustrations!!
battlefield	3	0.18	3/3 (100%)	its a battlefield and you are waging chemical war against the most determined opponent
duel	2	0.12	0/2	duel carriageway
enemy	97	5.78	9/20 (45%)	fear of the unknown is sometimes the enemy cancer is a deadly enemy you are a cruel wicked disease and you are a universal enemy no-one, no matter how close, ever got it!! that this was a war - we had won one battle, but we knew the enemy was regrouping we are our own worst enemy i wouldn't wish it on my worst enemy
struggle	698	41.56	17/20 (85%)	it's a struggle to motivate myself even getting out of a chair was a struggle
victory	24	1.43	2/20 (10%)	grasp the victory of a second chance of life reaching each milestone is a minor victory enjoying murray's victory

3.3.7 Section summary: investigation into metaphoricity of terms

The qualitative evaluation of potentially metaphoric terms in context has shown that the nouns *journey*, *step*, *move*, *rollercoaster*, *coaster*, *fight*, *struggle*, *fighter*, *battle*, and *battlefield*, are all used predominantly metaphorically in this context, while the other terms evaluated are not predominantly used metaphorically. In addition, examples of metaphor mixing were found, which may represent more deliberate use of metaphor.

The summary table below shows that the metaphoric terms *journey*, *step*, and *rollercoaster*, are more prevalent in the cancer.net.20 data (more established community members), than in the comparative cancer.net.LT20 data (less established community members). The occurrence of *fight* is approximately the same in both conditions, while *battle* is used less by more established participants.

Table 3-12 density of metaphoric terms for different segments of the cancer.net community

metaphoric term	cancer.net.20 density	cancer.net.LT20 density	cancer.net.20 / cancer.net.LT20
journey	332.23	162.73	2.04
rollercoaster	48.94	28.56	1.71
step	118.42	94.74	1.25
fight	61.45	59.79	1.03
battle	58.35	77.86	0.75

These results support the interpretation of *journey*, *step*, and *rollercoaster* as dominant metaphors within this community, with *journey* being the most prominent community metaphor, since it is used over twice as often by established participants compared to newer participants. In comparison *fight*, and *battle*, while still used predominantly metaphorically, may not be part of the characterisation of this community, since *fight* has the same density for both populations, and *battle* is used less by more established participants than by less established participants.

The nouns found to be metaphoric in at least 90% of uses are considered in subsequent analysis, grouped into themes based on the primary terms found via topic modelling and their associated metaphoric nouns found via the word vector similarity method. The following table summarises the metaphor themes and their component nouns.

Table 3-13 cancer.net community metaphor themes

metaphor theme	nouns that form the theme
journey	journey
rollercoaster	rollercoaster, coaster
fight	fight, fighter
battle	battle, battlefield

Although the majority of potentially metaphoric near synonyms of metaphoric terms, the terms found through the word vector similarity investigations in chapter two, were found to be used predominantly literally, the process of exploration around the original terms gives confidence that all related dominant metaphor terms for this community have been found. This is important for the analysis in future chapters, in which the relationship between metaphor use and language style is considered through the application of methods to compare language style between posts that do and do not contain particular identified metaphor themes.

In the next section graph-based ranking of metaphoric phrases, and comparison of the density of noun phrases in cancer.net.20 and cancer.net.LT20, are used to locate typical metaphoric phrases and metaphoremes in cancer.net. This provides more insight into whether a particular metaphor can be considered to be a community metaphor or not, i.e. a metaphor that is dominant in the community and has been adapted to do particular work within it. It also provides insights into what that work is.

3.4 Metaphoric phrases and metaphoremes

In this section to further explore the typical use of metaphor on cancer.net the prevalence of different metaphoric noun phrases relating to the community metaphor themes is evaluated. The presence of metaphoremes is also considered, where, as discussed in chapter one, a metaphoreme, which may be part of a wider metaphor theme, is a semantic and pragmatic unit of analysis that combines "the linguistic, the cognitive, the affective, and the socio-cultural" (Cameron and Deignan, 2006, p. 686) in a particular speech community at a particular time.

Graph-based ranking is used here to rank each metaphoric noun phrase based on the combined rank in the data as a whole of all nouns in that noun phrase. This supports the interpretation of phrases as more or less important depending on the typical position within a sentence of their component nouns, such that more central, and consequently more connected, nouns may be considered to be more important to the meaning of the corpus. A subsequent count-based analysis compares the density of noun phrases containing metaphoric terms in cancer.net.20 (more established participants) and cancer.net.LT20 (less established participants). Since the data from each participant group contains a different number of posts, in order that between-group comparisons can be made the percentage is calculated of each metaphoric phrase as a proportion of all included phrases relating to the specific metaphoric term under investigation. If dominant metaphoric phrases are more used by more established participants, this supports the interpretation of them as metaphoremes harnessed to carry particular social and psychological work in this community.

In all cases, in order to focus on more typical discourse, phrases are only listed where they constitute at least 2% of all included phrases and are present at least twice, while non-metaphoric phrases meeting these criteria are crossed out.

Where data is analysed based on the presence or absence of a particular metaphor theme, for each post in the data, for each noun phrase in that post, if that noun phrase contains a noun whose lemma is in the metaphor theme(s) under investigation, that post is identified as containing that metaphor theme. This identification method is discussed further in chapter four, where it is referred to as the NOUNS method.

3.4.1 Graph-based ranking introduction

As discussed in chapter one, it has been argued that the form of language reflects conceptual metaphorical meaning. For example, the conceptual metaphor CLOSENESS IS STRENGTH OF EFFECT influences understanding of literal closeness within forms in a sentence as reflecting the strength of the closeness of the relationship holding between those forms (Lakoff and Johnson, 2003). In previous analyses in chapter two, apart from the establishment of multi-word terms as n-grams, the order of terms in a document was not considered. Graph-based ranking in comparison is a method that takes account of the position of a term within its sentence, and its closeness to other terms.

In graph-based ranking a graph expresses the structure between variables. A graph in this context is made up of vertices, which are connected by edges. For the purposes of the current investigation the vertices are word lemmas, with the edges representing the relationships between them. In graph-based ranking:

1. when one vertex links to another vertex, it increases the importance of that other vertex

2. the importance of the linking vertex determines the importance of the links it creates

3.4.1.1 PageRank

Google PageRank (Brin and Page, 1998) is a well known example of a graph-based ranking algorithm which uses collective knowledge of the Web to rank web pages.

PageRank determines how important a web page is by counting the number and quality of links made to the page: it allows for connections to have a variable value, in that being linked to from a high-ranking page has more benefit in terms of rank than does being linked to from a lower-ranked page. The PageRank algorithm was so much more successful in finding relevant pages in comparison to previous methods that it changed how the Web is searched such that Google has become by far the most dominant search engine.

3.4.1.2 TextRank

Based on PageRank, TextRank is a graph-based ranking model for text, in which the importance of a particular vertex is determined by calculating global information recursively from the whole graph, in addition to the use of local vertex-specific information (Mihalcea and Tarau, 2004). The TextRank method can be summarised as follows:

1. Identify the lexical units that best define the task at hand, then add the relevant terms as vertices in the graph
2. Identify relations that connect the terms, and use those relations to draw edges between vertices in the graph
3. Iterate the graph-based ranking algorithm until convergence

4. Sort vertices based on their final score, then use the values attached to each vertex for ranking/selection decisions (Mihalcea and Tarau, 2004)

The PyTextRank Python library (Nathan *et al.*, 2021) is used in the current investigation to calculate the TextRank of noun phrases extracted from cancer.net.20. Before graph-based ranking is applied to that data, a short example is worked through, to support use of the method to identify highly ranked metaphoric phrases as metaphoremes.

3.4.2 Short example of graph-based ranking

The following excerpt from a randomly selected post is used for this example:

i had a little chuckle to myself when you described the problems with your medical compression suit. it was very brave of you to even try to wear the suit so soon after surgery.

The terms used for this example investigation are the lemmas of all adjectives, nouns, and verbs in this post excerpt. The rank is calculated for those lemmas, then those lemma ranks are used to calculate the rank of phrases extracted from the text. The actual lemmas included in the analysis, in the order they appear, are described in the following table.

Table 3-14 lemmas from each sentence included in graph-based ranking analysis

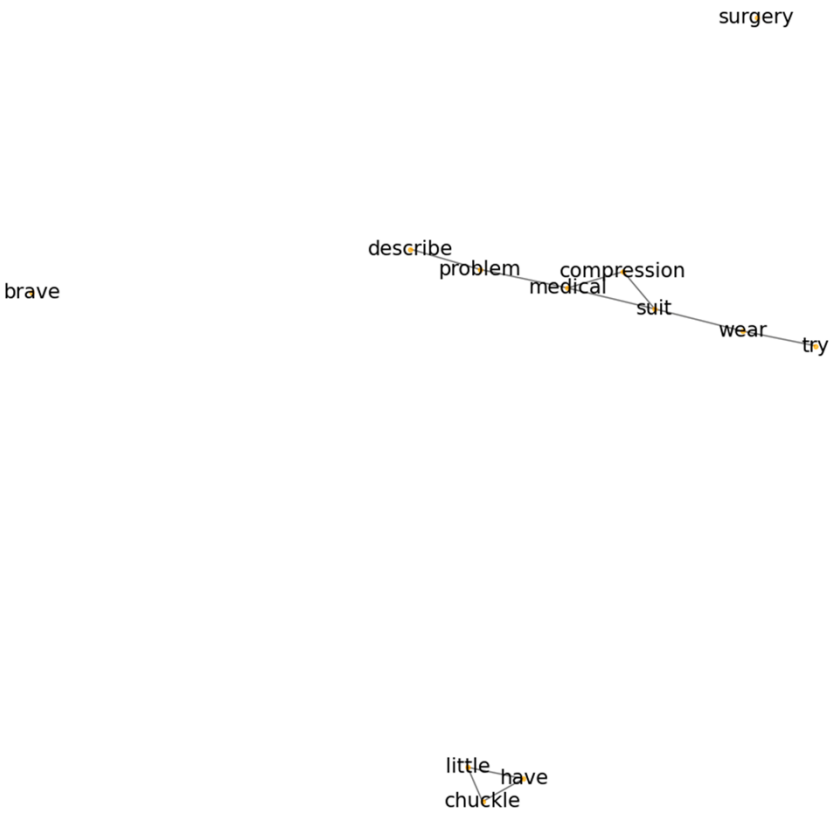
sentence	lemmas from sentence included in analysis
1	had little chuckle describe problem medical compression suit
2	brave try wear suit surgery

3.4.2.1 Construct a lemma graph of the text

The first step in the application of graph-based ranking to the text is to construct a lemma graph in which lemmas are connected to other lemmas co-occurring within a window of a maximum of N words, with PyTextRank setting N=3. It is in this way that the order of lemmas affects their rank, since lemmas in the middle of a sentence are more likely to be connected to other lemmas within that window of three lemmas behind and ahead. For

example, in the lemma graph below the lemmas *have*, *little*, and *chuckle*, are linked to each other since they appear within three words of each other in the first sentence. They are not linked to any other lemmas since all other lemmas are positioned more than three words away from them. The lemma *brave* is not linked to any other lemmas since in the original sentences it is positioned more than three words away from any other lemma. The lemma *suit* is linked to *medical* and *compression* through its location in the first sentence, and to *wear* through its location in the second sentence.

Figure 3-5 lemma graph of nouns, verbs, and adjectives



3.4.2.2 Calculating the PageRank score for each vertex in the lemma graph

After the lemma graph is constructed, the rank score for each vertex is set to 1, then the ranking algorithm is run for several iterations until it converges at a threshold of 0.0001 (Nathan *et al.*, 2021). The following table lists the calculated rank score for each lemma.

Table 3-15 lemmas and associated rank score

lemma	rank score
medical	0.149
suit	0.149
problem	0.113
wear	0.113
little	0.108
chuckle	0.108
compression	0.101
describe	0.064
try	0.064
brave	0.016
surgery	0.016

3.4.2.3 Calculating the PageRank score for each phrase in the text

For each noun phrase in the text a composite rank score is then calculated based on the rank scores of the component ranked lemmas of that phrase. The noun phrases from the example excerpt are listed in rank order in the table below.

noun phrase	rank score
your medical compression suit	0.20
a little chuckle	0.12
the suit	0.10
the problems	0.08
surgery	0.06
it	0.00

In this two-sentence example the highest ranked noun phrase is *your medical compression suit*. Since the two example sentences are about the addressee's medical compression suit, graph-based ranking can therefore be said to have located the most meaningful noun phrase in this short text.

3.4.3 Graph-based ranking of the cancer.net.20 data

In the following application of graph-based ranking to the cancer.net.20 data only noun lemmas were used to calculate the rank of individual noun phrases, because, as discussed

in chapter two, the focus in the current study is on commonly used metaphoric terms in this data as they are instantiated as nouns. In order to concentrate similar phrases based on the same underlying metaphor, each noun phrase was contracted to only its nominal components. For example, the noun phrases *my cancer journey* and *your cancer journey* were consolidated into the single phrase *cancer journey*.

3.4.4 All cancer.net.20 data

Graph-based ranking was applied to the cancer.net.20 data as a whole, such that the ten top ranked noun phrases in the following table can be understood as representing the dominant concerns of cancer.net.20, which it can be seen are literal and related to the specific forum issue of living with cancer. None of the community metaphor terms found previously are present in these phrases, which is to be expected since metaphor is relatively sparse in this data. While the phrases are listed in rank order, in fact there is little difference between the top rank score of 0.0676 and the tenth rank score of 0.06: these ten top ranked phrases, each of which contain the noun *cancer*, all have a similar relevance within cancer.net.20.

Table 3-16 top ranked phrases from cancer.net.20

	phrase	rank score	count
1.	cancer treatment	0.0676	535
2.	breast cancer	0.0671	6165
3.	breast cancer stage	0.0645	15
4.	breakthrough breast cancer	0.0623	5
5.	cancer diagnosis	0.0622	949
6.	lung cancer	0.0614	2178
7.	cancer radiotherapy	0.0610	5
8.	cancer chat	0.0603	1007
9.	throat cancer	0.0602	382
10.	cancer centre	0.0600	92

3.4.5 journey phrases

For the analysis in this section cancer.net.20 posts that contain the *journey* theme were extracted to consider the presence of *journey* metaphoremes, and to consider more closely the interpretative repertoires associated with use of *journey*.

3.4.5.1 journey noun phrases ranked

In the results table below, the 10 top ranked noun phrases from posts containing the *journey* theme are listed, regardless of whether they contain the actual *journey* term.

Table 3-17 ranked phrases from posts containing journey theme

	noun phrase	rank	count
1.	cancer journey	0.0925	818
2.	cancer treatment	0.0839	39
3.	*life cancer	0.0816	8
4.	mum cancer	0.0811	9
5.	breast cancer	0.0807	464
6.	breast cancer nurse	0.0784	6
7.	breast cancer diagnosis	0.0780	7
8.	cancer diagnosis	0.0770	104
9.	lung cancer	0.0758	78
10.	breast cancer lady	0.0757	7

It can be seen from the table above that apart from the highest ranked phrase *cancer journey*, all other phrases are literal and relate to the specific experience of living with cancer. This dominance of the metaphoric phrase *cancer journey* suggests that it is a metaphoreme in this context.

In addition, the person references *mum cancer*, *breast cancer nurse*, and *breast cancer lady*, are present in this list, whereas no person references are present in the top ranked phrases for the more general data of the previous section. This suggests that the *journey* metaphor may be used in conjunction with discussion of the personal experience of cancer. It is also notable that all of these personal references are female (*mum*, *lady*) or

relate to a predominantly female experience (breast cancer). The single time reference is *life cancer*, marked with an asterisk.

3.4.5.2 Noun phrases that contain journey ranked

For the analysis in this section, as in the previous section all cancer.net.20 posts were used that contain the *journey* metaphor theme, but only the noun phrases containing *journey* are listed in the table below. It can be seen that the potential metaphoreme *cancer journey* is the highest ranked *journey* phrase. In addition, the single time reference *life journey* supports the topic modelling insight that *journey* is used to frame the experience of living with cancer in the longer term, including a whole life.

Table 3-18 ranked journey phrases

	journey noun phrase	rank	count
1.	cancer journey	0.0780	818
2.	husband journey	0.0757	24
3.	*life journey	0.0716	34
4.	treatment journey	0.0689	15
5.	chemo journey	0.0682	9
6.	breast cancer journey	0.0676	15
7.	work journey	0.0676	2
8.	grief journey	0.0673	15
9.	car journey	0.0673	12
10.	return journey	0.0672	7

3.4.5.3 Between group comparison of journey noun phrases

In the following table a comparison of the prevalence of each *journey* noun phrase is made between the more established participants who have made 20 or more posts (cancer.net.20, on the left of the table), and the less established participants who have made fewer than 20 posts (cancer.net.LT20, on the right of the table), on the basis that metaphoremes used to do particular work in this community are more likely to be used by more established participants. To make the results more comparable between the groups,

the proportion of each *journey* phrase in terms of all included *journey* phrases was calculated. For all between group comparisons, in order to focus on more typical discourse, rather than idiosyncratic discourse, reported results are limited to those that occur as at least 2% of all included phrases, and occur at least twice.

Table 3-19 journey phrases in comparative participant groups based on number of posts made

	journey phrase cancer.net.20	count	%	journey phrase cancer.net.LT20	count	%
1.	cancer journey	818	65	cancer journey	84	32
2.	*year journey	51	4	hour journey	15	6
3.	*life journey	34	3	mum journey	15	6
4.	husband journey	24	2	*year journey	9	3
5.				*life journey	8	3
6.				grief journey	7	3
7.				people journey	5	2
8.				chemo journey	4	2
9.				*month journey	4	2
10.				mile journey	4	2
11.				husband journey	4	2
12.				recovery journey	4	2

The results table above shows that although *cancer journey* is the most prevalent *journey* noun phrase for both groups, it constitutes 65% of included phrases for the more established group, in comparison to 32% for the less established group. This strongly supports the previous interpretation that *cancer journey* is a metaphoreme in this context, such that it has a particular meaning and value in this community. In addition, *cancer journey* is so dominant in the more established group that there are far fewer qualifying *journey* phrases. For the more established participants the time phrases *year journey* and *life journey* are the next most prevalent, again supporting the insight from topic modelling in chapter two that *journey* is used in this context to frame longer periods of time, including the span of a whole life.

3.4.6 rollercoaster phrases

For the analysis in this section cancer.net.20 posts that contain the *rollercoaster* theme were extracted to consider more closely the interpretative repertoires associated with its use. In order to address the variety of spellings of rollercoaster, which are disruptive for this analysis, the strings 'roller coaster', 'roller-coaster', and 'roll coaster' were first replaced with the string 'rollercoaster'.

3.4.6.1 rollercoaster noun phrases ranked

In the first results table below, the 10 top ranked noun phrases from all posts containing the *rollercoaster* theme are listed, regardless of whether they contain nouns from the theme. Three metaphoric phrases are present, which are underlined in the results. The presence of the phrase *cancer journey*, which has a similar density of use to the phrase *cancer rollercoaster*, shows that *cancer journey* is used in conjunction with the *rollercoaster* theme, and this further supports its identity as a metaphoreme.

Table 3-20 ranked phrases from posts containing the rollercoaster metaphor theme

	noun phrase	rank	count
1.	listening ear	0.0746	5
2.	<u>cancer rollercoaster</u>	0.0720	39
3.	<u>cancer journey</u>	0.0691	35
4.	<u>cancer ride</u>	0.0690	2
5.	cancer diagnosis	0.0689	21
6.	word cancer	0.0686	6
7.	lung cancer	0.0685	15
8.	cancer patient	0.0685	9
9.	brain cancer	0.0684	2
10.	cancer type	0.0684	2

3.4.6.2 Noun phrases that contain coaster ranked

For the analysis in this section, only those metaphoric noun phrases containing nouns from the *rollercoaster* theme were extracted, and these are listed, by rank score, in the

table below. It can be seen that the metaphoric phrase *cancer rollercoaster* is the highest ranked. The time reference *life rollercoaster* is also present.

Table 3-21 ranked rollercoaster phrases

	rollercoaster phrase	rank	count
1.	cancer rollercoaster	0.0720	39
2.	word rollercoaster	0.0679	3
3.	grief rollercoaster	0.0677	4
4.	*life rollercoaster	0.0544	3

3.4.6.3 Between group comparison of rollercoaster noun phrases

For the analysis in this section, a comparison of the prevalence of each *rollercoaster* phrase was made between more established participants (on the left of the results table below), and less established participants (on the right of the table). Reported results are limited to those that occur as at least 2% of all included phrases and occur at least twice. It can be seen that *cancer rollercoaster* is the most prevalent *rollercoaster* phrase for both groups but is over 2.5 times as common in the discourse of more established in comparison to less established participants. The relatively low prevalence of the *rollercoaster* theme, particularly in the posts of less established participants, means that there are few phrases that meet the inclusion criteria.

Table 3-22 coaster phrases in comparative participant groups based on the number of posts made

	rollercoaster phrase	count	%	rollercoaster phrase	count	%
	cancer.net.20			cancer.net.LT20		
1.	cancer rollercoaster	39	53	cancer rollercoaster	2	20
2.	grief rollercoaster	4	5			
3.	*life rollercoaster	3	4			
4.	word rollercoaster	3	4			

3.4.7 fight phrases

For the analysis in this section posts in cancer.net.20 that contain the *fight* metaphor theme were extracted, so that the interpretative repertoires around its use and the presence of metaphoremes could be considered more closely.

3.4.7.1 fight noun phrases ranked

In the results table below, the 10 top ranked noun phrases from all posts containing the *fight* theme are listed, regardless of whether they contain nouns from the *fight* theme. The two metaphoric phrases are underlined, and the time reference *half year* is also present. It is notable that the metaphoreme *cancer journey* is not present in conjunction with the *fight* theme.

Table 3-23 ranked phrases from posts containing the fight theme

	noun phrase	rank	count
1.	head neck cancer	0.0828	2
2.	side effect	0.0798	48
3.	cancer patient	0.0777	14
4.	*half year	0.0750	5
5.	cancer tumor	0.0745	2
6.	listening ear	0.0728	3
7.	<u>cancer fight</u>	0.0716	4
8.	chemo nurse	0.0706	5
9.	<u>cancer fighter</u>	0.0693	4
10.	breast cancer	0.0691	72

3.4.7.2 noun phrases that contain fight ranked

For the analysis in this section, metaphoric noun phrases containing nouns from the *fight* theme were extracted, and these are listed according to rank score in the results table below. Metaphoric phrases associating cancer specifically with *fight* are the highest ranked, with the time reference *year fight* the next highest ranked.

Table 3-24 ranked fight phrases

noun phrase	rank	count
1. cancer fight	0.0716	4
2. cancer fighter	0.0693	4
3. *year fight	0.0606	7
4. bun fight	0.0563	3
5. snowball fight	0.0562	2
6. breast cancer fighter	0.0557	3
7. *month fight	0.0541	2
8. mum fight	0.0464	3

3.4.7.3 Between group comparison of fight noun phrase proportions

In the following table, a comparison of the prevalence of each metaphoric phrase is made between more established participants (on the left of the table), and less established participants (on the right of the table). It can be seen that *year fight* and *cancer fighter* are the most prevalent *fight* noun phrases for both participant groups.

Table 3-25 fight phrases in comparative participant groups based on the number of posts made

fight phrase cancer.net.20	count	%	fight phrase cancer.net.LT20	count	%
1. *year fight	7	15	*year fight	12	21
2. cancer fighter	4	9	cancer fighter	4	7
3. cancer fight	4	9	mum fight	4	7
4. breast cancer fighter	3	7	wife fight	3	5
5. bun fight	3	7	cancer fight	3	5
6. mum fight	3	7	water fight	2	3
7. *month fight	2	4	*month fight	2	3
8. snowball fight	2	4			

3.4.8 battle phrases

For the analysis in this section posts in cancer.net.20 that contain the *battle* metaphor theme were extracted, to consider more closely the interpretative repertoires and potential metaphoremes around its use.

3.4.8.1 battle noun phrases ranked

In the results table below, the 10 top ranked noun phrases from all posts containing the *battle* theme are listed, regardless of whether they contain nouns from the *battle* theme. The highest ranked phrase is *half year battle*, and the metaphoric phrases *cancer battle* and *cancer journey* are also present, showing that these metaphors are used together; this again supports the view that *cancer journey* is a metaphoreme.

Table 3-26 ranked phrases from posts containing the battle theme

	noun phrase	rank	count
1.	*half year battle	0.0815	3
2.	cancer tumor	0.0745	2
3.	cancer battle	0.0728	31
4.	cancer journey	0.0693	23
5.	word cancer	0.0693	4
6.	cancer part	0.0692	2
7.	cancer diagnosis	0.0692	9
8.	cancer cell	0.0689	4
9.	cancer centre	0.0689	2
10.	prostate cancer	0.0689	17

3.4.8.2 Noun phrases that contain battle ranked

For the analysis in this section the same data was used, but only the noun phrases containing a noun from the *battle* theme were extracted and these are listed by rank score in the table below. The time references *half year battle*, *year battle*, *month battle*, *week battle*, and *yr battle* are present (a concordance analysis confirmed that *yr battle* means *year battle* rather than *your battle*), suggesting that the *battle* theme is strongly associated with discussion of time.

Table 3-27 ranked phrases containing battle

	battle noun phrase	rank	count
1.	*half year battle	0.0815	3
2.	cancer battle	0.0728	31
3.	*year battle	0.0660	45
4.	*month battle	0.0648	20
5.	health battle	0.0579	2
6.	*week battle	0.0558	9
7.	*yr battle	0.0543	4
8.	dad battle	0.0484	5
9.	husband battle	0.0460	2
10.	mum battle	0.0426	6

3.4.8.3 Between group comparison of battle noun phrase proportions

In the following table, a comparison of the prevalence of each metaphoric phrase is made between more established participants (on the left of the table), and less established participants (on the right of the table). It can be seen that the time reference *year battle* is the most prevalent *battle* noun phrase for both more established and less established participants, and there are multiple time references for both groups, marked with an asterisk. The phrases are very similar for both groups, however, *cancer battle* is 2.5 times as prevalent in the more established group, while *month battle* is around twice as prevalent in the less established group.

Table 3-28 battle phrases in comparative groups based on the number of posts made

	battle phrase	count	%		battle phrase	count	%
	cancer.net.20				cancer.net.LT20		
1.	*year battle	45	29		*year battle	49	36
2.	cancer battle	31	20		*month battle	33	24
3.	*month battle	20	13		cancer battle	11	8
4.	*week battle	9	6		*week battle	8	6
5.	mum battle	6	4		mum battle	4	3
6.	dad battle	5	3		*day battle	3	2
7.	*yr battle	4	3		dad battle	3	2
8.	*half year battle	3	2		*half yr battle	3	2

3.4.9 Section summary: metaphoric phrases and metaphoremes

In this section, in order to explore typical cancer.net metaphoric noun phrases and potential metaphoremes, graph-based ranking was used to order noun phrases based on the ranking of their component noun lemmas in the cancer.net.20 data as a whole, such that nouns that are typically more central in any sentence have a higher rank. This was done on the basis that more connected, more central, noun phrases are more likely to represent the meaning and purpose of a discourse.

In addition, for each metaphor theme a count-based comparison between the discourse of more established and less established participants on cancer.net provided insights into the different work of the metaphor themes in this community. Although specific metaphoric phrases were discovered, no new metaphoric terms were identified, again suggesting that all dominant metaphor terms in cancer.net have been located.

The analysis in this section found that *cancer journey* is by far the most dominant metaphoric *journey* phrase, accounting for 65% of *journey* phrases made by more established participants and 32% of *journey* phrases made by less established participants. It is also used in conjunction with the *rollercoaster* and *battle* themes, but not with the *fight* theme. These results strongly support the view that *cancer journey* is a metaphoreme in this context and that it may support an ideological dilemma. Similarly, *cancer rollercoaster* is the most dominant phrase for the rollercoaster theme, accounting for 53% of *rollercoaster* phrases for more established participants and 20% of *rollercoaster* phrases for less established participants. For the fight theme the most dominant phrase is *year fight*, and similarly for the battle theme it is *year battle*. There were time references associated with each metaphor theme, which are summarised in the following tables.

Table 3-29 density of time phrases for comparative participant groups for each metaphor theme

journey time phrase	% cancer.net.20	% cancer.net.LT20
year journey	4	3
life journey	3	3
month journey		2
TOTAL	7	8

rollercoaster time phrase	% cancer.net.20	cancer.net.LT20
life rollercoaster	4	
TOTAL	4	

fight time phrase	% cancer.net.20	% cancer.net.LT20
year fight	15	21
month fight	4	3
TOTAL	19	24

battle time phrase	% cancer.net.20	% cancer.net.LT20
year battle	29	36
month battle	13	24
week battle	6	6
yr battle	3	
half year battle	2	
day battle		2
half yr battle		2
TOTAL	53	70

It can be seen from the tables above that while there are comparatively fewer time references for the *journey* and *rollercoaster* themes, which it has been shown are more prevalent in the discourse of more established participants, for the *fight* theme, which is equally prevalent for both participant groups, around one fifth of the total noun phrases include the time references *year fight*, or *month fight*, with *year fight* by far the most used of these.

Results for the *battle* theme, which is used more by more recent than by more established participants, suggest that *battle* is predominantly used to frame the experience of cancer in terms of periods of time, including *day*, *week*, *month*, *half-year*, and *year*. For the *battle* theme time references account for over half of all *battle* phrases for the more established group, and 70% of all *battle* phrases for the less established group. It is also notable that the time reference *life* is used in conjunction with both *journey* and *rollercoaster*, but not in conjunction with *fight* or *battle*.

3.5 Signalled metaphor

The signalling devices investigated in this section include similes, copular similes, a sort/kind of, feel as if/though, and explicit markers such as *metaphor*, and *analogy*. The use of such co-textual markers of metaphor has been suggested as "a partial solution to the vexed problem of locating Active metaphors in a corpus" (Goatly, 2011, p183). For example, a corpus study found text written for explanatory or teaching purposes to contain more metaphor signals than were found in text written to be spoken. A large majority of the signals were copular similes, verbal processes, and modals/conditionals. In addition, conditional signals (*if...could*, and *as it were*) were found to always foreshadow metaphorically expressed possibilities (Skorczynska and Ahrens, 2015).

As discussed previously, MIPVU identifies metaphor markers such as *like*, *as*, *as if*, and *so-called* as related to potentially deliberate metaphor, but excludes general signals of indirectness, including *sort of* and *kind of*. Cameron and Deignan (2003) argue that although an analyst might describe such terms as vague, they are often associated with an interpersonal perspective in that they serve to reduce the directness of an assertion, such that it may have exactly the amount of precision that the speaker intended. This is consistent with Zhang's (2011, p. 573) characterisation of such vague language as "a

linguistic unit (word, phrase or utterance) that has an unspecified meaning boundary, so that its interpretation is elastic in the sense that it can be stretched or shrunk according to the strategic needs of communication". Cameron and Deignan (2003) recognise general signals such as *sort of* and *kind of* as tuning devices, which flag potential ambiguity in the interpretation of text, and suggest what sort of interpretation is intended. They can also be used to locate metaphor. For example, in their investigation a small, intensively studied corpus yielded the following tuning devices which were then explored in a larger corpus: *actually, almost, imagine, just, kind of, a little, really, sort of, and as it were*. It was found that such tuning devices are often used together, and that the tuning device *sort of* frequently signals potentially deliberate metaphor. And further, given the frequency of the use of tuning devices with metaphor, it was suggested that the absence of a tuning device may indicate that direct and unmitigated metaphor is being used (Cameron and Deignan, 2003). In an analysis of extended metaphors from different psychotherapist-patient pairs, Tay (2011a) shows that the signalling of metaphor may also extend beyond the concept of tuning the degree of metaphoricity of a statement, to more global discourse objectives, for example structuring and drawing attention to a stepwise progression of extended metaphors. The examples of discourse markers found by Tay (2011a) include *sort of* and *kind of*, as well as more general markers that may not be so specifically associated with metaphor, such as *you know, I mean, and you know what*. For these reasons, the signalling devices *a sort of* and *a kind of* are considered in the current investigation. Consideration is also given to the relationship between signalled metaphor and the characteristic metaphor themes of this community.

In this section, since the focus is on potentially more idiosyncratic metaphor rather than dominant community metaphor themes, all of the cancer.net data is used, not just the data for participants who have made 20 or more posts.

3.5.1 Using pattern matching to identify metaphor

Linguistic pattern matching based on hand-compiled linguistic reference sources such as WordNet has been used as a way to try to locate metaphor automatically in text. For example, Krishnakumaran and Zhu (2007) focused on the metaphoric effect of verbs and adjectives on nouns, which they classified as the following types:

Table 3-30 types of linguistic metaphor

type	format	example	premise for detection
I	target noun copula verb source noun	He is a brave lion	the lower the relatedness between the source and target nouns, the greater the likelihood that the phrase is metaphorical
II	a verb (not copula) acts on a noun	He planted good ideas in their minds	incongruous verb-noun pairs are identified by a violation of selectional preference
III	an adjective acts on a noun	He has a fertile imagination	incongruous adjective-noun pairs and a shift from abstract to concrete visualization

For the identification of type I metaphors, the WordNet hyponym relation was used such that a phrase was classified as metaphoric if there was no hyponym relation between the subject and object of that phrase. For identifying metaphor types II and III the WordNet hyponym relation was used in conjunction with bigram co-occurrence information from a Web IT corpus. Based on a test set derived from the Master Metaphor List, an identification accuracy of 58% was reported with this method.

In another linguistic approach to identifying metaphor automatically in text Wilks *et al.* (2013) used a semantic parser to first identify the semantic role of verbs, and then abstract those verbs to their highest-level hypernyms in WordNet. If the primary sense of a verb does not satisfy the primary preference of the context, but a lower (less frequent) sense does, then use of the verb in that context was considered likely to be metaphorical. Problems included, for example, that the primary WordNet sense of the verb *erase*, which

is "remove from memory or existence", is metaphorical, while the literal sense is ranked second. An F1 accuracy score of 0.67 was reported with this method (Wilks *et al.*, 2013).

Although hand-constructed resources such as WordNet have been used in many methods designed to find metaphor automatically in text (Shutova, 2015; Leong *et al.*, 2020) it has been suggested that they do not have the density and diversity of connections needed to support realistic automatic interpretation of metaphor (Veale, Shutova and Klebanov, 2016). As discussed in chapter two, this has led more recently to a focus on transformer-based deep learning models to identify metaphor in large text corpora.

3.5.2 Hearst patterns

Hearst patterns (Hearst, 1992) are easily-recognisable lexico-syntactic patterns that occur frequently, and across text genre boundaries, that reliably indicate a lexical relationship of interest without the need for pre-encoded knowledge such as WordNet. The establishment of Hearst patterns was motivated by the need for applicability to a wide range of texts. In comparison to statistical language models used to locate metaphor, which require a statistically significant number of relationships, Hearst patterns require only one instance of a relationship. In comparison to the research examples in the previous section, which use WordNet to find hyponymy relations, the original Hearst patterns were based on "the automatic acquisition of the hyponymy lexical relation from unrestricted text" (Hearst, 1992, p. 539). The following Hearst pattern examples are from the original paper:

1. X such as Y
2. X as Y
3. X and/or Y
4. X especially Y

In a wide-ranging investigation into computational approaches to metaphor the Hearst patterns listed in the following table were considered as methods for identifying figurative language automatically in text (Veale, Shutova and Klebanov, 2016).

P=single adjectival property

N=single noun

V=single verb

Table 3-31 Hearst patterns considered by Veale et al. for locating metaphor automatically

metaphor type	Hearst pattern	examples given
simile patterns	as [P] as a [N]	as cute as a baby
	[V+ing] like a [N]	crying like a baby
copular metaphors	X is Y	tobacco is poison
	X is a Y	Kate Moss is a pencil
conceptual metaphors	Xs are Ys	argument is war
		time is money

Although the copular metaphor and conceptual metaphor Hearst patterns suggested by Veale *et al.* (2016) were not found in the current investigation to be productive in locating metaphor in the complex, naturally occurring, cancer.net data, the simile patterns were found to be productive, and are investigated further below.

3.5.3 spaCy matcher

The spaCy matcher is a rule-matching engine that operates over tokens. Compared to the use of regular expressions on raw text, the matcher also gives access to the tokens within documents, and the relationships between them. The matcher rules can refer to token annotations (for example a token's `text` or `tag_`), and flags (for example `IS_PUNCT`). In the current investigation the spaCy matcher is used to find lexico-syntactic patterns representing signalled metaphor, such as the Hearst patterns for similes described above. No pre-encoded knowledge, such as WordNet hyponyms or synonyms, is used with this method.

3.5.4 Similes

The spaCy matcher patterns in the following table were used to find similes in the Hearst pattern formats described above by Veale *et al.* (2016).

Table 3-32 matcher patterns to find similes

simile pattern	matcher pattern
1. as [P] as a [N]	{LEMMA:as}, {POS:ADJ}, {LEMMA:as}, {LOWER:{IN: [a, an]}}, {POS: NOUN}
2. [V+ing] like a [N]	{TAG:VBG}, {LEMMA: like}}, {LOWER:{IN: [a, an]}}, {POS: NOUN}

3.5.4.1 Pattern 1 similes: as [P] as a [N]

The 191 pattern 1 similes found in cancer.net are typically commonly used metaphoric comparisons that are a well-established general language resource. Since there are only 191 pattern 1 similes in total, they are not very prevalent in this large corpus of 260,241 forum posts. In the following table, all pattern 1 similes are listed that occur more than once.

Table 3-33 pattern 1 similes occurring more than once in cancer.net

pattern 1 simile	count
1. as fit as a fiddle	13
2. change is as good as a rest	12
3. as sharp as a pin	10
4. as stubborn as a mule	8
5. as bright as a button	8
6. as flat as a pancake	5
7. as sick as a dog	5
8. as weak as a kitten	4
9. as high as a kite	4
10. as white as a sheet	3
11. as fit as a flea	3
12. as strong as an ox	3
13. as stubborn as an ox	2
14. as brown as a berry	2
15. as sharp as a tack	2
16. as soft as a baby	2

3.5.4.2 Pattern 2 similes: [V+ing] like a [N]

There are 574 instances of pattern 2 similes in cancer.net, which are predominantly commonly used comparisons such as *sleeping like a log*, *sobbing like a baby*, *working like a dog*, and *eating like a horse*. However, there are also idiosyncratic instances of this participle verb-based simile that are evocative of the lived experience of their speaker, some of which are listed in the table below.

Table 3-34 idiosyncratic pattern 2 similes from cancer.net
pattern 2 simile

1. i do worry about looking like a car crash with multiple scars
2. feeling like a pincushion
3. looking like a lobster
4. looking like a lop-sided robin redbreast
5. looking like a hedgehog
6. he's lying in bed looking like a bag of bones

3.5.4.3 Simile patterns in different metaphoric conditions

To consider the relationship between use of these simile patterns and use of community metaphor, in this section five metaphoric conditions were investigated, that consist of all posts that contain the relevant metaphor theme. The no metaphor themes condition (5) was matched in size to the *journey* condition (7,164 posts) for comparison. The following table summarises the instances of each simile pattern for each metaphoric condition. The number of posts per pattern instance is shown in brackets so that the density of occurrence in the different conditions can be compared.

Table 3-35 number of posts in each metaphoric condition that contain simile patterns

	posts contain	posts	pattern 1	pattern 2	combined
1.	journey theme	7,164	10 (716)	10 (716)	20 (358)
2.	rollercoaster theme	1,507	4 (377)	3 (502)	7 (215)
3.	fight theme	1,945	0	4 (486)	4 (486)
4.	battle theme	2,004	3 (668)	5 (401)	8 (251)
5.	no metaphor theme	7,164	3 (2,388)	5 (1,433)	8 (896)

The chi-square test of independence was used to test the significance of the difference between the metaphoric conditions for this and all subsequent metaphor signalling analysis, and effect size was calculated using Cramér's V. In conjunction with p values, effect size is an important consideration since it takes into account sample size, which is of particular relevance when sample sizes are large, which may lead to high χ^2 values and very low p values, as is the case in some analysis in the current section. In terms of the independence of the conditions, Brezina (2018, pp. 112–113) notes that since a corpus is "a sample of texts which combine a number of interconnected linguistic features", "with corpora we need to relax this assumption". While there may be some relationship between use of the different metaphor themes, the data used is very diverse, from many participants and discussion threads, and covers a long period of time, such that it is unlikely there is a consistent relationship between the themes. It can be seen from the following table that there is little overlap in use of the different themes, for example 93% of uses of the *journey* theme are isolated from the other themes. In addition, the primary concern in the current analysis is the distinction between use of different signalling methods in metaphoric posts compared to non-metaphoric posts.

Table 3-36 distribution and overlap of metaphor themes

metaphor theme	posts	% of all posts	exclusive posts	% exclusive posts	overlap posts
journey	7,164	2.75	6,643	93	521
rollercoaster	1,507	0.58	1,233	82	274
fight	1,945	0.75	1,660	85	285
battle	2,004	0.77	1,693	84	311
any theme	11,906	4.57			

Since it is a typical rule of using chi-square that there should be at least five of each observed and expected frequency, the values from the two simile patterns were combined, and the *fight* condition was excluded. It was found that there is a significant difference between the metaphoric conditions $\chi^2(0)=647$, $p<0.001$, Cramér's $V=0.27$, showing that use of these simile patterns varies with the presence or absence of the metaphor themes investigated, and that they are least prevalent where none of the metaphor themes is present in a post.

3.5.5 Copular similes

For the analysis in this section the following matcher pattern was used to find copular similes in the cancer.net data:

```
{POS:NOUN}, {LEMMA[be]}, {LEMMA:like}, {LOWER:{IN: [a, an]}}, {OP: ?}, {POS:NOUN}
```

162 instances of copular similes were found in cancer.net, with the six copular similes listed in the table below each occurring more than once. These repeated copular similes account for 21 occurrences altogether, with the phrase *life is like a book* used seven times. However, *life is like a book* is used by the same participant on each occasion, across a time period from the start of May 2018 to September 2019: this metaphor is not adopted by any other participant - it does not become part of the interpretative repertoire on

cancer.net. And although the participant who uses this metaphor continues to be active on cancer.net throughout the rest of the data, they also do not use the phrase after September 2019.

**Table 3-37 copular similes that occur more than once
repeated copular simile examples**

	count
1. life is like a book	7
2. cancer is like a rollercoaster	4
3. emotions are like a rollercoaster	4
4. life is like a box of chocolates	2
5. recovery is like a walk of 1000 miles	2
6. that first year is like a wound that won't heal	2

Although Goatly (2011) suggests that similes kill metaphor, the remaining 141 copular similes (87% of all the copular similes) each occur only once, are largely creative, and are not typically commonly used phrases. And further, although it has been argued that most novel metaphors are creative extensions of existing conceptual mappings (Lakoff and Johnson, 2003), the creative examples listed below do not just extend typical metaphor in a new way, rather they are seemingly deeply felt personal expressions of the lived experience of the participant. This use of creative metaphor may be related to a study showing that while stress typically involves the expression of more conventional metaphor, anxiety is more likely to be associated with use of more creative metaphor, an effect which is explained by the interpretation of stress as a comparatively more general, everyday occurrence than anxiety (Moskaluk, 2020).

**Table 3-38 creative examples of copular similes
copular simile examples from cancer.net**

1. my head is like an electricity pylon with all the thoughts and worries
2. cancer is like a computer programme that runs in the background but that you don't notice while you use other programmes and then at times it simply puts up a notification or pop-up window that interrupts the other programs
3. i am like a petulant child in a toy shop being offered every teddy there is to replace the one i lost
4. his body is like an oil filled radiator filled with chemotherapy and radiotherapy and it has to slowly leech out vicious circle as water food and of all things

5. a woman getting it is like a unicorn turning up out of the blue
6. the afterlife i think is like a swimming pool
7. it is like a spider's web and is impossible to remove all the tiny threads which is why it is so difficult to treat
8. a woman is like a tea bag - you don't know how strong she is until you put her in hot water
9. it is like a roundabout with little chance of getting off
10. fear is like a nasty little worm in your ear telling you nasty things to make your life miserable & the only way to get rid of it is to rid yourself of the fear
11. information on one's cancer is like an onion
12. this disease is like a thief in the night
13. its like a tongue coming back to a sore tooth
14. grief is like a spiral staircase
15. the relatives do want those they love to undergo treatment even when hope is like a fading candle
16. depression is like a nasty thing biting at your heels when it senses a weakness
17. my head is like a potato sprouting cress
18. it is like a tornado in my head and the only way to get away from it is to open up a trap door in my mind and crawl inside it
19. the mind is like a 3-D home entertainment system and you are the witness or the viewer of the mind
20. i was told that chemo in the body is like a bath filling up with water week by week
21. life is like a book many chapters and your dad filled a lot of your mums story
22. its like a comfy blanket of knowing they are not alone
23. its like a deep cut that takes ages to heal which it does but always leaves a scar
24. my husband is like a patchwork quilt with all his operations
25. the vein in my hand is like an undercooked piece of spaghetti
26. chemo is like a face lift
27. my head was like a bowling ball
28. love from your family is like an additional medicine
29. the waiting time is like a ticking bomb
30. crying is like a safety valve
31. emotions/grief are like a battlefield aren't they
32. my phone was like a christmas tree with all the emails from you two

3.5.5.1 Copular similes in different metaphoric conditions

To consider the relationship between use of copular similes and use of the established metaphor themes, for the analysis in this section five metaphoric conditions were investigated, that are defined by the presence or absence of the metaphor themes. The no metaphor theme condition (5) was matched in size to the *journey* condition (7,164 posts)

for comparison. The ratio is the number of posts in each condition divided by the number of copular similes, such that a lower ratio value represents a higher density of copular similes.

Table 3-39 number of posts in each metaphoric condition that contain copular similes

	posts contain	posts	count	ratio
1.	journey theme	7,164	22	325
2.	rollercoaster theme	1,507	16	94
3.	fight theme	1,945	5	389
4.	battle theme	2,004	4	501
5.	no metaphor theme	7,164	5	1,433

The chi-square test of independence shows that there is a significant difference between the conditions $\chi^2(0)=1943$, $p<0.001$, Cramér's $V=0.42$. Copular similes, which are idiosyncratic and expressive in 87% of uses, are significantly more likely to be used in posts in which the metaphor themes are present, in comparison to posts in which none of the themes are present.

3.5.6 Clausal similes: as if / as though

The clausal simile terms *as if* and *as though* were considered, with the following spaCy matcher pattern used to find metaphor signalled by these terms:

{LEMMA: as}, {LOWER: {IN:[if, though]}}

It was found that the phrases *as if* (1,406 instances), and *as though* (3,047 instances), are prevalent in the data, with *as though* occurring over twice as often as *as if*. A sample of 10 of each phrase type was extracted and these are listed in the tables below.

Table 3-40 examples of as if

as if examples (from 1,406 instances)

1.	its as if i'm saying well i'm not going anywhere
2.	i was amazed at how many people said to me well, he was a good age when i told them as if it was okay
3.	i feel as if i am going out of my mind.

4. i am down and feel so badly as if i could have done more
5. when she was diagnosed with an incurable brain tumour i felt as if my heart would break
6. that made me feel as if they have given up on him
7. some people don't show emotion and love to the ones they should do as if it is a weakness it is very sad
8. when i asked if she did self examination she looked at me as if i had 2 heads
9. i have always felt a fraud when i have been in there, as if i am taking a much-needed hospital bed
10. how can i put on a brave face and act as if nothing is happening

**Table 3-41 examples of as though
as though examples (from 3,047 instances)**

1.	because i look fine they carry on as though i am fine
2.	it felt as though my skin was dying
3.	i feel as though i've been frozen
4.	i feel as though Ill feel like this forever
5.	i know it is still early days, but i feel as though my heart is breaking
6.	i feel as though i am failing to do my job as a mother
7.	it has just come as such a big shock i feel as though i'm dreaming
8.	told me that even though he had felt as though he was dying the day before, he felt well enough now to get up on the roof and do an hours painting
9.	don't feel as though i am losing my mind now
10.	it sounds as though you are doing okay though and getting the help that you need

3.5.6.1 as if/ as though in different metaphoric conditions

To consider the relationship between use of *as if* and *as though* and use of community metaphor, five metaphoric conditions were investigated, that are defined by the presence or absence of the established metaphor themes. The no metaphor condition (5) was matched in size to the *journey* condition (7,164 posts) for comparison. The number of posts per pattern is shown in brackets.

Table 3-42 density of as if/as though in different metaphoric conditions

posts contain	posts	as if	as though	combined
1. journey theme	7,164	165 (43)	69 (104)	234 (31)
2. rollercoaster theme	1,507	37 (41)	18 (84)	55 (27)
3. fight theme	1,945	51 (38)	16 (122)	67 (29)
4. battle battle	2,004	51 (39)	27 (74)	78 (26)
5. no metaphor theme	7,164	87 (82)	36 (199)	123 (58)

The chi-square test of independence shows that the difference between the conditions is significant ($p < 0.001$), such that the metaphor signalling devices *as if* and *as though* are more common in posts in which the identified metaphor themes are used than in posts where they are not used. This effect is strongest for *as though*, and is least strong when *as if* and *as though* are considered together, showing that they are used differently.

Table 3-43 significance of difference between metaphoric themes for as if and as though signalling devices

signalling device	χ^2	df	p	Cramér's V
as if	29	0	<0.001	0.17
as though	85	0	<0.001	0.19
as if and as though combined	21	0	<0.001	0.18

3.5.6.2 feel as if/ feel as though in different metaphoric conditions

Qualitative consideration of *as if* and *as though* phrases explored in the previous section showed that the most evocative of these phrases tend to be associated with the verb *feel*, for example:

Table 3-44 examples of feel as if/feel as though

feel as if/feel as though examples
1. i felt as though you came right round to my house and picked me up by the scruff of the neck and said "right come on you get on board"
2. it just feels as if the world had gone away
3. feels as though i'm walking through mud
4. now i feel as if someone has cut out the centre of my body and i am hollow

5. feel as if my whole world has come crashing down around me
6. i feel as though i'm on hold until i find out if its spread, sort of fight wise
7. i felt as if i should just slip back into life
8. i just felt as if my world had caved in and was near hysterical at times
9. i feel as if it is someone else that has gone as if i have been to someone else's funeral
10. i feel as though my heart is being ripped out very slowly bit by bit

Since there is a relatively high prevalence of the *as if* and *as though* signalling devices, the matcher was amended to apply to the verb *feel*:

{LEMMA:feel},{LEMMA:as},{LOWER:{IN:[if,though]}}

The count of occurrences of each signalling device is displayed in the following table, with the number of posts per signalling device shown in brackets.

Table 3-45 density of feel as if/feel as though in different metaphoric conditions

posts contain	posts	feel as if	feel as though	combined
1. journey theme	7,164	34 (211)	18 (398)	52 (138)
2. rollercoaster theme	1,507	11 (137)	9 (167)	20 (75)
3. fight theme	1,945	10 (194)	7 (278)	17 (114)
4. battle theme	2,004	8 (250)	13 (154)	21 (95)
5. no metaphor theme	7,164	20 (358)	11 (651)	31 (231)

The chi-square test of independence shows that the differences between the conditions is significant ($p < 0.001$) for both *feel as if* and *feel as though*, but it is strongest for *feel as though*. These signalling devices are more likely to be used in posts in which the identified metaphor themes are present than in posts in which they are not present. In both cases the effect is stronger than when *as if* and *as though* are considered more generally, in the previous section. So, the signalling devices *feel as if* and *feel as though* are more

likely to co-occur with identified community metaphor themes than are the signalling devices *as if* and *as though*.

Table 3-46 significance of between metaphoric themes for as if and as though signalling devices

signalling device	χ^2	df	p	Cramér's V
feel as if	118	0	<0.001	0.16
feel as though	509	0	<0.001	0.28
feel as if and feel as though combined	113	0	<0.001	0.21

3.5.7 so to speak/as it were

The following matcher patterns were used to locate the signalling devices *so to speak* and *as it were*:

{LOWER:so},{LOWER:to},{LOWER:speak}

{LOWER:as},{LOWER:it},{LOWER:were}

There are 322 instances of *so to speak*, which typically signals more self-conscious use of relatively conventional metaphor, such as *back to normal*, *too much on my plate*, *music to my ears*, *start the ball rolling* etc. Examples are listed in the table below.

Table 3-47 examples of so to speak
so to speak examples

1.	i'm not in a boat on my own so to speak
2.	it all seems back to normal so to speak
3.	feels like a bit too much on my plate so to speak
4.	that to me is music to my ears, so to speak
5.	putting your feelings down on paper so to' speak
6.	start the ball rolling so to speak
7.	i will leave the ball in their court so to speak
8.	i have found talking to strangers helps as you don't have to spare their feelings so to speak
9.	just to test the water so to speak
10.	were going to hear her fate so to speak

There are 50 instances of *as it were*, and these also typically signal relatively conventional metaphor.

**Table 3-48 examples of as it were
as it were examples**

1.	i'm trying to go with the flow as it were
2.	i have been on the treatment treadmill as it were
3.	needs must as it were
4.	outside the circle, as it were
5.	have a chance to draw a line as it were
6.	get a reduction on the other one to make the same size as it were
7.	i can pick up things between the lines as it were
8.	when its all done and dusted as it were
9.	i know i am going down that road and entering the end as it were
10.	i'm an outdoor cat as it were

The count of occurrences of each signalling device is displayed in the following table, with the number of posts per signalling device shown in brackets.

Table 3-49 occurrence of so to speak/as it were in different metaphoric conditions

	posts contain	posts	so to speak	as it were	combined
1.	journey theme	7,164	16 (448)	5 (1,433)	21 (341)
2.	rollercoaster theme	1,507	1 (1,507)	1 (1,507)	2 (754)
3.	fight theme	1,945	6 (324)	3 (648)	9 (216)
4.	battle theme	2,004	6 (334)	1 (2,004)	7 (286)
5.	no metaphor theme	7,164	6 (1,194)	0	6 (1,194)

Since it is a typical rule of using chi-square to test independence that all of the observed and expected values are >5, the combined values were considered for significant differences, and data for the *rollercoaster* condition was excluded. The difference between the included conditions is significant $\chi^2(0)=1242$, $p<0.001$, Cramér's $V=0.45$, showing that the signalling devices *so to speak* and *as it were* are more likely to occur in

posts where the identified metaphor themes (excluding *rollercoaster*) are present, than in posts where none of the identified themes are present.

3.5.8 Superordinate terms: a sort of/a kind of

Although the superordinate terms *a kind of* and *a sort of* are not included in MIPVU, as discussed above such 'tuning devices' have been found to be used in conjunction with metaphor, and have been associated with potentially deliberate metaphor (Cameron and Deignan, 2003). The following matcher pattern was used to find them in cancer.net:

{LEMMA: a},{LEMMA: {IN:[sort,kind]}},{LEMMA: of}

There are 155 instances of the term *a sort of*, and 144 instances of the term *a kind of*.

Although the signalling of metaphors appears to express the intention of their writer that they are metaphoric, whether they are understood as metaphoric is also dependent on the reader, and this seems to be more the case precisely because of the distancing effect of the superordinate term.

**Table 3-50 examples of superordinate terms
a sort of/a kind of examples**

1.	she was diagnosed with cancer about a year after my wife and she looked on her as a kind of talisman
2.	our city is hit by floods and it has a sort of wet carnival atmosphere
3.	i have woken up panicking about brain tumours before and sent myself in to a sort of frenzy
4.	she said her consultant told her that chemo was a kind of extra insurance policy which was really a great way to look at it my advice to you
5.	how much solace you have given me, its like a kind of cyber love
6.	don't think no one cares because we do as we have or are going through loss you're in a sort of club now no one wants to be in
7.	for me seeing them all gave comfort and a sort of closure to the life part
8.	but its a kind of staging post in the grief process & lets you move on to the next stage
9.	waiting for appointments & test results is a kind of living nightmare which feels impossible to get out of
10.	there is a sort of unspoken embarrassed apology that they don't really know what to say to you

The count of occurrences of each signalling term in different metaphor conditions is displayed in the following table, with the number of posts per instance of the signalling term shown in brackets.

Table 3-51 density of a sort of/a kind of as though in different metaphoric conditions

posts contain	posts included	a sort of	a kind of	combined
1. journey theme	7,164	9 (796)	11 (651)	20 (358)
2. rollercoaster theme	1,507	3 (502)	4 (377)	7 (215)
3. fight theme	1,945	6 (324)	1 (1,945)	7 (278)
4. battle theme	2,004	3 (668)	3 (668)	6 (334)
5. no metaphor theme	7,164	2 (3,582)	2 (3,582)	4 (1,791)

Since it is a typical rule of using chi-square to test independence that all of the observed and expected frequencies should be at least 5, the difference between the metaphoric conditions was considered for the combined constructions only; the no themes condition was included even though the number of combined occurrences is only 4, because it is the specific condition of focus. The difference between the metaphoric conditions is significant $\chi^2(0)=3023$, $p<0.001$, Cramér's $V=0.5$, showing that these superordinate terms are more likely to be used in posts where the community metaphor themes are also used, than in posts in which they are not used. The difference between the metaphoric conditions also is significant, when the no themes condition is excluded, although the difference is much weaker $\chi^2(0)=41$, $p<0.001$, Cramér's $V=0.11$.

Vague language such as *a sort of/a kind of* has been associated with the discussion of sensitive topics, such that as the level of sensitivity of a topic increases, so the level of vagueness in talk-in-interactions also increases (Zhang, 2013). The increased use of these

vague terms in conjunction with identified metaphor themes, then, may indicate that those metaphor themes may support the discussion of sensitive topics in this community.

3.5.9 Explicit markers

For the analysis in this section, explicit markers of metaphor were considered, including the lemmas: *metaphor*, *figurative*, *trope*, *analogy*, *metonymy*, *allegory*, and *personification*. The following spaCy matcher pattern was used to extract all sentences in which these explicit markers occur:

```
{LEMMA: {IN:[metaphor,figurative,trope,analogy,metonymy,allegory,personification]}}
```

There are 57 unique sentences that contain explicit markers of metaphor, of which 12 include the lemma *metaphor*, and 45 include the lemma *analogy*. Apart from one instance (example 1), all of these sentences include metaphor, much of it creative, and some include extended metaphor (13, 27), and mixed metaphor (16).

Table 3-52 examples containing 'metaphor'
examples of the explicit marker 'metaphor'

1.	it had no metaphors i didn't understand
2.	to try to make this a bit more concrete, let me try a metaphor ... its not great, but its as far as i have got so far
3.	i sometimes use the metaphor that this is like trying to stop dandelions spreading in a garden
4.	apparently, i bled like a stuck pig, although i don't think that's the correct metaphor
5.	a lot of time and effort taken writing in metaphors and then when your goddess is told you are sick she just up and bolts with your son?
6.	most of this language of metaphors really annoys me especially when spoken in a caring headmistress to a child sort of way
7.	i tend to use garden metaphors - the doctors are pulling out the big weed which has unfortunately taken hold-doing this will leave a hole which they will fill with weed killer but doing nothing means the garden will be quickly swamped
8.	i'm in the mode of seeing this as half-time following the sporting metaphor
9.	sometimes we have to jump up and down and shout quite loudly metaphorically to be heard by doctors
10.	i use the tread mill metaphor one step at a time and you can't stop keep going
11.	things look strangely different from this end of the metaphorical stethoscope.

12. i will give myself a metaphorical slap and get a grip

Table 3-53 examples containing 'analogy'
examples of the explicit marker 'analogy'

13.	my counsellor uses an analogy whereby the "hole" in my life will continue and i will move forward with it always there but very gradually i will build round the hole (a bit like a tree can grow around a hole)
14.	so i've used the analogy of an army and explained that the bug is hiding behind the army that protects my body and the medicine has to knock the army over to get to the bug but then allows the army to get up again before it goes in for another battle
15.	your analogy of being hit by a steam train upon diagnosis is a pretty accurate description of how most of us feel
16.	that really shook me and i think for the first time ever i appreciated that every tool in the box is being used to make sure as you say in your lovely analogy that these muddy footprints are removed, and you will be sparkly clean very soon
17.	your analogy of it being a mental form of self harming sounds very true
18.	your analogy of it hitting you in the face with a shovel is so accurate
19.	so totally get the ticking time-bomb analogy
20.	were also getting a new washing machine plumbed in that day, and i can't help but see the humour in this analogy!
21.	love the puzzle analogy!
22.	thinking back to my analogy of the pea in the straw that's what you have got - all still contained within the duct - none of the little blighters escaped and became invasive
23.	my analogy since this news was that i was in a big boat of cancer survivors, but now i'm being towed along in a little dingy behind them
24.	i love your military analogy - and i hope your scan results are positive
25.	your analogy of it hitting you in the face with a shovel is so accurate
26.	try **s analogy of pink boxing gloves at the ready, with all of us cancer pilgrims backing your corner
27.	i've used the analogy of an army, and explained that the bug is hiding behind the army that protects my body, and the medicine has to knock the army over to get to the bug, but then allows the army to get up again before it goes in for another battle

3.5.9.1 Explicit markers in different metaphoric conditions

To consider the relationship between use of explicit markers and use of the identified metaphor themes, in this section the five metaphoric conditions were investigated in conjunction with use of explicit markers. The no metaphor themes condition (5) was matched in size to the *journey* condition (7,164 posts) for comparison.

Table 3-54 occurrence of explicit markers in different metaphoric conditions

posts contain	posts included	metaphor count	analogy count
1. journey theme	7,164	0	4
2. rollercoaster theme	1,507	0	3
3. fight theme	1,945	0	1
4. battle theme	2,004	0	1
5. no metaphor theme	7,164	0	2

It can be seen from the results table that the signalling device *metaphor* is not used in conjunction with any of the community metaphor themes, and nor is it present in the random sample from the large number of non-metaphoric posts. For *analogy* also there are not enough instances to apply the chi-square test of independence, to consider whether the difference between the metaphoric conditions is significant. Although these signalling devices do signal metaphor that seems to be used self-consciously, they are not notably associated with use of any of the identified cancer.net metaphor themes.

3.5.10 Section summary: signalled metaphor

Using pattern matching to find metaphor signalling devices has been productive. Each such pattern found was sampled and checked in its context and was found to predominantly signal use of metaphor. In the table below, the number of posts per metaphor signalling device is summarised for each metaphoric condition considered above, such that a lower number represents a higher density of metaphor in that condition, in that there are fewer posts per instance of a signalling device. The entries have been ordered by effect size of the difference between the metaphoric conditions, such that the first entry, a sort of/a kind of, shows the most significant difference between conditions.

Table 3-55 density of metaphor signalling type by metaphoric condition, ordered by effect size

signalling pattern	journey 1	rollercoaster 2	fight 3	battle 4	none 5	χ^2	Cramér's V
i. a sort of a kind of	358	215	278	334	1,791	3,023	0.50
ii. copular simile	325	94	389	501	1,443	1,943	0.42
iii. so to speak as it were	341	754	216	286	1,194	1,242	0.45
iv. as [P] as a [N] [V+ing] like a [N]	358	215	486	251	896	647	0.27
v. feel as though	398	167	278	154	651	509	0.28
vi. as though	104	84	122	74	199	85	0.19
vii. as if	43	41	38	39	82	29	0.17
viii. feel as if	211	137	194	250	358	118	0.16
AVERAGE	267	213	250	236	827		

It can be seen from the table above that the density of each metaphor signalling pattern is lowest for the condition in which no community metaphor themes are present (condition 5). When the average density of all patterns for each theme is considered it can be seen that the results are similar for each metaphoric condition, while for the non-metaphoric condition the presence of signalled metaphor is over three times less dense.

For each metaphor signalling pattern listed there is a significant difference between the metaphoric conditions ($p < 0.001$), for which the effect is strongest for the vague language *a sort of* and *a kind of* that is specifically not included in MIPVU. The increased prevalence of *a sort of* and *a kind of* alongside identified metaphor themes suggests that although they were found to signal relatively conventional metaphor, these superordinate terms do have an additional relationship to metaphor. So, while very common metaphorical phrases may be dismissed as noise, "their frequency, and their affective value both suggest that they are of importance and should be placed at the centre of investigation rather than on the periphery" (Cameron and Deignan, 2006, p. 689). This

also supports the view of metaphor as vague language that may be used to target and realise communication goals (Zhang, 2011), with vague language also more likely to be used when sensitive topics are discussed (Zhang, 2013). While the

The second strongest effect was for copular similes, which typically signalled creative metaphor, showing that signalled creative metaphor is more likely to occur in posts in which dominant metaphor themes of this community are used, than in posts in which they are not used. The identification of copular similes as the most creative form of signalled metaphor is also consistent with the career of metaphor theory discussed in chapter one, which identifies novel metaphors as employing between-domain mapping based on comparison. In the career of metaphor, similes are found to invite a comparison of the target domain to the source domain (comparison), while metaphors promote the classification of the target domain as a member of a category named by the source domain (categorization) (Bowdle and Gentner, 2005).

3.6 Chapter discussion

Qualitative analysis was used to identify which potentially metaphoric terms are actually metaphoric in the cancer.net context. Of the primary metaphoric terms found via topic modelling it was found that *journey* and *rollercoaster* are more prevalent in data for participants who have made 20 or more posts; *battle* is more prevalent in data for participants who have made fewer than 20 posts; and *fight* is equally prevalent for both sets of participants. This suggests that *journey* and *rollercoaster* may be community metaphors on cancer.net. But it is also important to consider their relationship to the *fight* and *battle* themes, for example, do they suppress use of those potentially non-community themes. This is explored further in the diachronic analysis in chapter five.

Table 3-56 density of metaphoric terms for different segments of the cancer.net community

metaphoric term	cancer.net.20 density	cancer.net.LT20 density	cancer.net.20/ cancer.net.LT20
journey	332.23	162.73	2.04
rollercoaster	48.94	28.56	1.71
fight	61.45	59.79	1.03
battle	58.35	77.86	0.75

In terms of the work of these metaphors in this context, the phrase *cancer journey* was located as a metaphoreme which may be used to represent a particular framing and evaluation of the experience of cancer. Time references used in conjunction with the themes show that *journey* may be used to frame the experience of cancer in the longer term as a *year journey*, and in the context of a whole life, as a *life journey*, while the *rollercoaster* theme also frames life as a *life rollercoaster*. However, for both *journey* and *rollercoaster* time references are not dominant.

For the *fight* and *battle* themes in comparison their time references do not include *life*; for *fight* they include *year*, and *month*; while for *battle* they include *day*, *week*, *month*, *half year*, and *year*, and account for over half of all *battle* phrases for both participant groups. The *battle* theme, then, is predominantly used to consider the experience of cancer in terms of periods of time, but not in the longer-term perspective of a whole life.

Metaphor signalling devices were found to occur more frequently in posts where the community metaphor themes are also present, compared to posts in which they are not present. In addition, this effect was found to be very strong for copular similes, which in this context are typically less conventional, more creative, metaphor. This suggests that use of the metaphor themes encourages increased use of other figurative language that is expressive beyond the interpretative repertoires associated in this community with those metaphor themes. The identified metaphor themes may represent more enduring metaphorical mappings with established entailments in this context, while the expression

of new experience or insights may require more creative metaphor. For example it has been shown that people use metaphor more when describing more intense emotional experiences than when describing actions (Fainsilber and Ortony, 1987). In addition creative metaphor has been found to be significantly more likely than conventional metaphor to perform an evaluative function (Fuoli, Littlemore and Turner, 2021). The identification of copular similes as the most creative form of signalled metaphor is also consistent with the career of metaphor theory described in chapter one, which identifies novel metaphors as employing between-domain mapping based on comparison.

The use of the signalling devices *a sort of* and *a kind of* although they are specifically excluded in MIPVU, were found to show the strongest effect between metaphoric conditions, being significantly more prevalent in posts in which the metaphor themes are used, compared to posts in which they are not used. This supports their relevance where metaphor use is under investigation, and particularly in terms of how the vagueness of metaphor allows it to be adapted for different purposes.

The intention set out in chapter one was to find metaphor themes in the data, where a metaphor theme is "an abstraction from the metaphorical statements in which it does or might occur. A metaphor-theme is available for repeated use, adaptation and modification by a variety of speakers or thinkers on any number of specific occasions" (Black, 1977, p. 438). In fact, rather than more general metaphor themes, individual metaphoric terms, and linguistically very closely related terms, have been found. These are still referred to as themes in subsequent analysis, and the following definition of the metaphor themes is used:

Table 3-57 cancer.net community metaphor themes

metaphor theme	nouns that form the theme
journey	journey
coaster	rollercoaster, coaster
fight	fight, fighter
battle	battle, battlefield

4 METAPHOR AND LANGUAGE STYLE

4.1 Introduction

In chapter three the cancer.net metaphor themes were established, and associated metaphoremes were considered. The highly ranked and prevalent metaphoric phrase *cancer journey* was identified as a metaphoreme which was found to be prominent when posts containing the *journey*, *rollercoaster*, or *battle* themes were selected, but not when posts containing *fight* were selected. The phrase *cancer rollercoaster* was also identified as a potential metaphoreme, although use was constrained to the condition in which the term *coaster* is present in a post. The work of *journey* was found to include the framing of the experience of cancer in the longer term of a *year* or *life*, and *battle* was found to be used to consider the experience of cancer over shorter time periods, as well as a *year*, and to be predominantly used to consider time.

Metaphor found via metaphor signalling devices was found to predominantly co-occur with posts in which the metaphor themes are present, in comparison to posts in which they are not present, suggesting that part of the work of the metaphor themes in this context includes the facilitation of the personal expression of how the difficult experience of living with cancer feels, including creative metaphor such as copular similes. The effect was strongest for the vague language constructions *a sort of* and *a kind of*, which may support the evolution of the work of the metaphor themes within this context. In addition, while these superordinate terms are not included in MIPVU (section 3.2.2), this shows that they have a relationship to metaphor.

In the current chapter, supervised learning, a type of statistical learning that is well suited to the large and diverse data of the current investigation, is used to consider the difference in language style between posts that do and do not contain the established metaphor

themes. Supervised learning here is supported by the previous unsupervised learning methods used to locate dominant characteristic metaphoric terms in cancer.net. Now that those metaphoric terms have been identified, that knowledge can be used to label cancer.net posts as containing or not containing the particular metaphor themes, which are summarised in table 4-1 below. A supervised algorithm is then trained on the labelled data supplied to it to predict the presence of the metaphor themes in each post. This technique is used here as a method of validating the organisation of the identified metaphoric terms into themes, on the basis that if an algorithm can be trained to accurately predict the separate themes, then they are distinct in terms of surrounding language style, and have been correctly identified in terms of their constituent terms. In the same way this method may support the focus on metaphor instantiated as nouns for the purposes of the current investigation (section 2.6.3.4). Supervised learning applied in this manner may also be used as a method to identify metaphor in a large (unlabelled) text corpus, and this also is explored in the current chapter.

Two supervised learning methods are compared. In the first, text-based, method, whole post texts are supplied to the algorithm. In this case the text is vectorized such that, unlike for the clustering, dimensionality reduction, and topic modelling investigations of chapter two, information about the order of terms in a post is retained. The algorithm in this case receives the text as it was written and can consider the connections between terms and their positioning within posts, as well as the semantic relationships between terms. In the second, count-based, method, the counts of 143 linguistic variables in each post are supplied to the algorithm, expressed as a proportion of the total word count. For example, the proportion of the first-person singular pronoun *i*; the proportion of identified *Anger* words; the proportion of possessive pronouns. This count-based method is more directly relatable to the LIWC approach discussed in chapter one (section 1.4). This comparison

of two very different methods of representing the text data, and their associated different supervised learning algorithms, also provides insights into the effect of identified metaphor on surrounding language style. Where the accuracy of the count-based method is higher this would suggest that it is the proportions of different linguistic variables that more accurately represent the work of the metaphor theme under consideration. Where the accuracy of the text-based method is higher, this would suggest that a focus on the positioning of terms within a text, and the relationships between terms, including semantic similarity, more accurately represents the work of the metaphor theme under consideration.

While the focus so far has been on metaphoric nouns, in order to further support that choice, in this chapter consideration is also given to the effect on surrounding language style of metaphor themes instantiated as nouns, in comparison to metaphor themes instantiated as both nouns and verbs. To support this investigation, the prevalence in cancer.net of each metaphoric term as nouns and verbs is summarised in the table below. It can be seen from the table that *fight* is over three times as often instantiated as a verb than as a noun, *battle* is instantiated around half as much as a verb than as a noun, and the consideration of metaphoric terms as a verb is not relevant to the *journey* and *rollercoaster* themes which are overwhelmingly, or completely, instantiated as nouns.

Table 4-1 summary of metaphor themes

metaphor theme	associated metaphoric terms	instances as a noun	instances as a verb	total instances
journey	journey	8,133	72	8,205
rollercoaster	rollercoaster	1,207	0	1,759
	coaster	552	0	
fight	fight	1,894	6,356	8,734
	fighter	484	0	
battle	battle	2,169	1,027	3,199
	battlefield	3	0	

Supervised learning is also used to consider the metaphor themes combined together, and separately, to further explore, for example, whether it makes more sense in terms of the work they do to group *journey* and *rollercoaster* into one theme rather than to consider them as separate themes.

Subsequently to the supervised learning analysis, a more detailed comparison is made of language style between posts that do and do contain a particular metaphor theme. This more detailed comparison provides further insights into the social and psychological work of the dominant metaphor themes in cancer.net, in terms of key attributes of language style associated with their use, which are related to state of mind. Cancer.net data is used in this chapter, not cancer.net.20, since the process of prospecting for typical metaphors of this discourse community by considering only the posts of established cancer.net participants is now complete. Cancer.net consists of all 260,239 posts collected from the forum, regardless of the number of posts participants have made.

4.2 Supervised learning

While in classical programming specific program rules to process data are written based on what is already known, a machine learning system in comparison is not explicitly programmed - it is trained to look for statistical structure, usually in large and complex datasets, such as cancer.net, for which mathematical statistics are not practical.

Supervised learning is the passing of labelled inputs (data plus answers) to a machine learning algorithm. That algorithm then works to produce a set of rules that allows the prediction of outputs (answers) from unlabelled input data (Chollet, 2018, p. 5). Whereas in chapter two unsupervised learning was used to find structure in unlabelled data, specifically to look for dominant metaphor in the cancer.net.20 data, in the current chapter supervised learning is used to consider whether there is a discernible difference in

language style between posts labelled as containing a metaphor theme, and posts labelled as not containing that metaphor theme.

4.2.1 Overview of the process

To implement supervised learning, each post in cancer.net was labelled according to whether the metaphor theme under consideration is or is not present in that post. The data was then split randomly into training, and test, datasets. Supervised learning was then run on the labelled training data, to produce a model to predict the category of each post in the set of test data. The predicted category for each post in the test data was then compared with its original (ground truth) label, in order to produce an accuracy score for the model. That accuracy score was used as a way of validating the established definitions of the metaphor themes, in that if good accuracy is obtained from the model, as well as demonstrating that each metaphor theme is surrounded by a distinctive (predictable) language style, it also supports the earlier methods used to define the themes: it demonstrates that community metaphor themes, and the language used to express them, have been identified accurately.

4.2.2 Supervised methods compared

Two supervised learning methods are explored in this chapter. The first, text-based, method takes as its input the text of each post. The second, count-based, method, takes as input the counts of 143 lexical categories in each post, for example the number of nouns, first-person singular pronouns, determiners, anger words, and so on. Both of these supervised methods must focus on the language style surrounding metaphor use to learn to predict the presence of the metaphor themes in a post, but each does this in a different way. As well as considering parts of speech and other linguistic attributes of terms, the text-based model also takes into account the word order of the text. The count-based

model in comparison focuses on the proportions of different linguistic variables in the text.

4.2.3 Evaluating the models

In machine learning the F1 score, which takes into account the support (number of documents) for each labelled condition, is the standard performance measure for a model.

F1 is the weighted average of precision and recall, where precision is the number of true positives predicted by a model divided by all positives predicted, and recall is the number of true positives predicted divided by the number of all positives

(*sklearn.metrics.f1_score* — *scikit-learn 0.24.2 documentation*, 2021).

Table 4-2 definition of F1 score

$$F1 = 2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall})$$

In the current study the scikit-learn Python library *classification_report* and *F1_score* are used to calculate the F1 score, with analytic focus on the weighted average F1, which accounts for any imbalance between labelled conditions, in terms of the number of posts in each condition. The F1 macro average in comparison does not take label imbalance into account: it is an average of the F1 score for the different conditions investigated.

While precision and recall are considered separately to support some methodological decisions of the current study, because for the purposes of the current investigation there is no reason to prioritise minimising either false positives or false negatives, the focus here is on the F1 weighted average score, which is a more general consideration of results that takes into account both precision and recall.

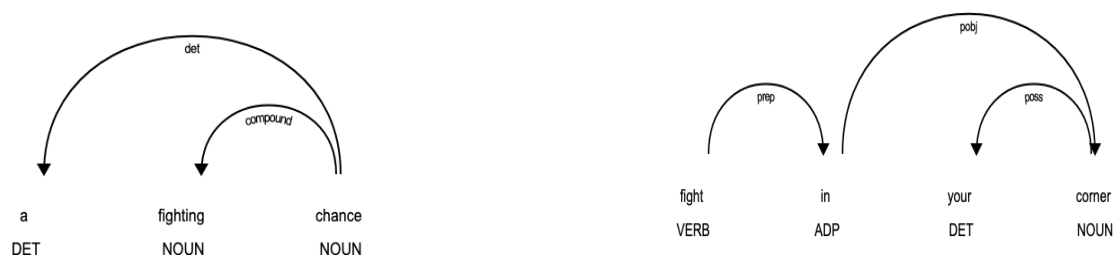
4.2.4 Two different labelling methods

Two different labelling methods are used in this investigation, which are based on the presence of metaphoric terms as nouns, or as nouns and verbs. This distinction is made in order to explore the hierarchy of force of different metaphor source terms, as discussed in chapter two, and to support the choice in the current investigation to focus on metaphors instantiated as nouns.

4.2.4.1 Labelling method one: nouns in noun phrases (NOUNS)

In labelling method one, a noun-based method, for each post in the data, for each noun phrase in that post, if that noun phrase contains a noun whose lemma is in the metaphor theme(s) under investigation, that post is labelled 1. Any posts that do not meet this condition are labelled 0. In the examples below, the noun phrase *a fighting chance* (left) would be counted as metaphoric, while the noun phrase *fight in your corner* (right) would not be counted as metaphoric, since *fight* is present as a verb, not a noun. This labelling method will be referred to as NOUNS.

Figure 4-1 noun phrases comparing metaphoric term as a noun, and as a verb



4.2.4.2 Labelling method two: nouns and verbs (NOUNS_VERBS)

In labelling method two, for each post in the data, if that post contains a noun or verb whose lemma is in the metaphor theme(s) under investigation, that post is labelled 1. Any

posts that do not meet that condition are labelled 0. In the examples used in the previous section, both the noun phrases *a fighting chance*, and *fight in your corner*, would be counted as metaphoric with this method, which will be referred to as NOUNS_VERBS.

4.2.5 Considering the number of posts in each condition

A further consideration is the matching of conditions in terms of number of posts supplied to the algorithm. There are many more posts that do not contain the metaphor themes than posts that do contain the themes, such that this affects the focus of the supervised algorithm in analysis in which the conditions are not matched in size. In the following investigations matching of condition sizes is achieved by extracting all of the metaphoric posts for a condition, and then making a random selection of the same number from the remaining posts. Although this leads to fewer posts being available for training and testing, because it forces the algorithm to pay more equal attention to the content of both metaphoric and non-metaphoric posts it may achieve better results than the situation in which condition sizes are not matched.

4.2.6 Section summary: about supervised learning

Supervised learning is the passing of labelled inputs to train an algorithm to produce a model to predict the labels of a set of unlabelled data. Supervised learning is evaluated using the F1 score, which takes into account the support for each condition being investigated. In the current investigation, two different supervised methods are explored. The first method (TEXT) passes to the supervised algorithm the whole text of individual posts, while the second method (COUNT) passes to the algorithm the counts of 143 different linguistic variables in each post. In addition, in this investigation two methods for labelling the posts supplied to the algorithm are considered, which are based on the instantiation of metaphoric terms as nouns (condition=NOUNS), and as nouns and verbs

(condition=NOUNS_VERBS). Finally, since cancer.net contains many more non-metaphoric than metaphoric posts, consideration is given to the matching of condition sizes for the data passed to the algorithm.

4.3 Text-based method (TEXT)

For the text-based method used in this section, the data passed to the supervised algorithm consists of the whole text of each labelled post. The text is vectorised such that the order of words is taken into account: this is different to the bag of words approach used previously with unsupervised learning.

4.3.1 Keras deep learning Python library

For this investigation the Keras deep learning Python library was used to set up a deep learning model, where deep means that successive layers of increasingly meaningful representations of the data are used, and the depth of a model represents the number of layers in that model (Chollet, 2018). Keras was designed to be accessible to non-specialists, to widen the areas of application of deep learning, for example into the social sciences, arts, and humanities.

Keras is an API designed for human beings, not machines. Keras follows best practices for reducing cognitive load: it offers consistent & simple APIs, it minimizes the number of user actions required for common use cases, and it provides clear & actionable error messages. It also has extensive documentation and developer guides. (Keras: the Python deep learning API, 2021)

On the Kaggle internet platform for data science competitions, in 2020 Keras was by far the most-used software tool for the top-5 teams, being used in 40 out of 120 top-5 teams, with the next nearest tool being used in 30 out of those 120 top-5 teams (*Keras: the Python deep learning API*, 2021).

4.3.2 Text-based model

In supervised learning, text is typically processed by recurrent layers with feedback connections. For example, LSTM (Long short-term memory) layers are recurrent neural network (RNN) layers that, unlike simple RNN layers, can carry information across many timesteps. Bidirectional LSTM layers are used in the current investigation to present the same information to the network in different ways, which can increase accuracy, and the memory of information across timesteps (Chollet, 2018). A single dropout layer is introduced to prevent over-fitting of the model to the training data, to make it more robust in terms of predicting the condition of a set of test data, where a dropout layer randomly sets to 0 a specified proportion of input units, in this case 20%.

Table 4-3 description of model, from Keras

Layer (type)	Output	Shape			Param #
embedding_1	(Embedding)	(None	None	32)	320000
bidirectional_1	(Bidirection	(None	None	64)	16640
dropout_1	(Dropout)	(None	None	64)	0
bidirectional_2	(Bidirection	(None	64)		24832
dense_1	(Dense)	(None	1)		65

The supervised algorithm is trained and validated over five epochs before being used to predict the label of each post in the test data, where an epoch is the number times that the algorithm works through the entire training dataset.

The final (output) layer is a sigmoid layer, which outputs a probability value between 0 and 1 which in the terms of the current investigation indicates how likely a particular post is to be metaphoric (condition 1) (Keras: the Python deep learning API, 2021). In order that the model can be evaluated, the output for the current investigation is classified as condition 1 where the sigmoid output is > 0.5 ; otherwise it is classified as condition 0. Although it is not pursued here, there is further investigative potential in focusing on the posts predicted with a higher or lower likelihood of being in each condition.

4.3.3 Comparison of different methods

4.3.3.1 Labelling method one (NOUNS), all metaphor themes combined

In this analysis, the single metaphoric condition *metaphors* is explored, such that the posts were labelled as containing metaphor, condition 1, if they contain a noun phrase containing a noun that is present in any of the metaphor themes. The model was run for unmatched and matched condition sizes (number of posts), and for conditions in which the metaphoric term(s) were retained in the text data (retain) or were removed from the data before it was passed to the supervised algorithm (remove). In the results table below test posts refers to the number of test posts in each condition, i.e. the non-metaphoric condition 0 or the metaphoric condition 1.

Table 4-4 prediction of posts containing any metaphoric term

metaphoric condition	remove/retain	match sizes?	weighted average F1	macro average F1	test posts 0	test posts 1
metaphors	retain	False	0.98	0.87	9,523	477
metaphors	retain	True	0.97	0.97	3,961	3,977
metaphors	remove	False	0.94	0.54	9,578	422
metaphors	remove	True	0.75	0.75	3,886	4,052

It can be seen from the results table above that the weighted average F1 is higher (the model is more accurate) when the metaphoric terms, for example *journey*, *fight*, are retained in the text data passed to the algorithm than when they are first removed from that data. This is to be expected, since a lot of information is being passed to the algorithm by the term retained in the data. Retaining or removing the term creates a different focus for the algorithm, and results in a difference in the posts predicted as containing or not containing the metaphoric term(s) of focus. When the metaphoric terms are removed from the data supplied to the algorithm, the focus of the algorithm is necessarily on the language surrounding the metaphoric term, rather than simply looking for different instantiations of the metaphoric term. For this reason, in the following

investigations only the *remove* condition is investigated, in which metaphoric terms are first removed from the data supplied to the algorithm.

Similarly, although the weighted average F1 is higher when condition size (number of posts in metaphoric and non-metaphoric conditions) is not matched, this may partly be related to the fact that since there are many more non-metaphoric than metaphoric posts, predicting a post as non-metaphoric is much more likely to be true. For example, the macro average F1, which does not take into account the number of posts in each condition, is much lower when condition sizes are not matched.

4.3.3.2 More detailed consideration of matched or unmatched condition sizes

In this section more detailed statistics are considered to further explore the difference in results when condition sizes are and are not matched, for the composite *metaphors* condition explored above. Since as discussed above retaining the metaphoric terms in the data passed to the algorithm is not productive in terms of the aims of the current investigation, only conditions in which those metaphoric terms are first removed from the data are considered in this section and going forward.

4.3.3.2.1 *unmatched condition sizes*

The following table provides a more detailed analysis of the previous result for unmatched condition sizes, in which there are many more posts labelled as non-metaphoric than posts labelled as metaphoric. Precision is the number of true positives predicted divided by all positives predicted. Recall is the number of true positives predicted divided by the number of all positives.

The low recall of 0.11 for class 1 (metaphoric) shows that the model is failing to identify true metaphoric posts; conversely the high recall of 0.97 for class 0 (non-metaphoric)

shows that the majority of posts labelled as non-metaphoric were identified as such. The numbers of posts true and predicted in each condition are described in the confusion matrix and visualised in the subsequent scatter plot. The disparity in precision and recall between the conditions illustrates a problem with such very unbalanced condition sizes, in which in this example nearly 96% of posts have been labelled as non-metaphoric: the focus on non-metaphoric posts leaves metaphoric posts less predictable. If all the posts were predicted to be non-metaphoric, in the current case that would give a high accuracy of nearly 96%.

Table 4-5 classification report for unmatched condition sizes, metaphoric terms first removed from the text

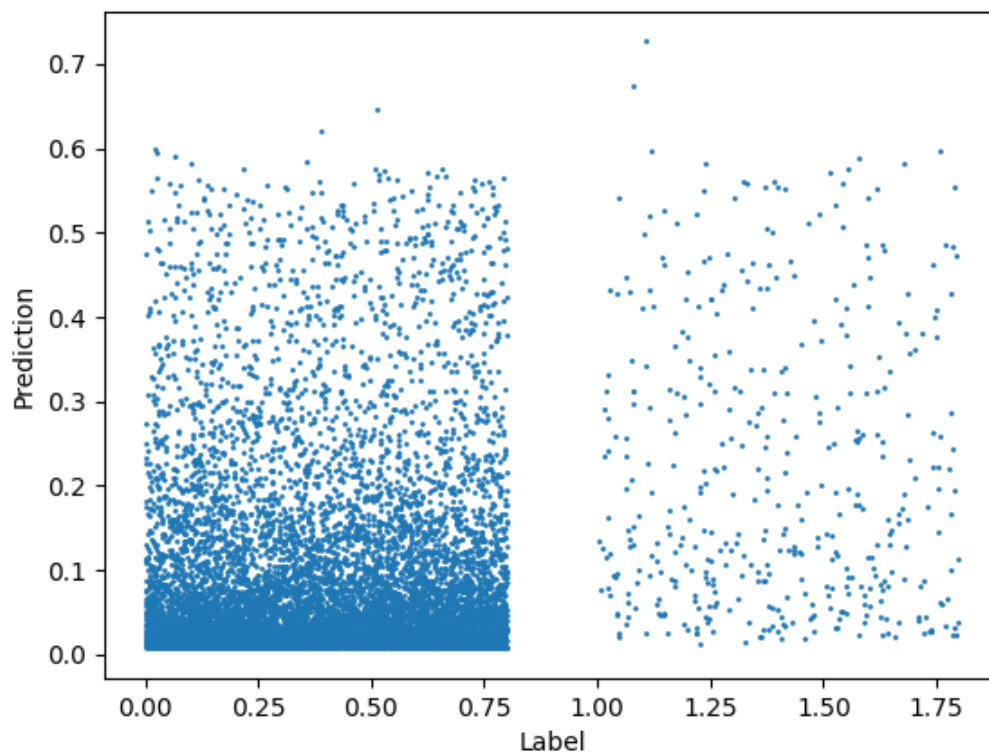
	precision	recall	f1-score	support
0	0.96	0.99	0.97	9578
1	0.19	0.08	0.11	422
accuracy			0.95	10000
macro avg	0.57	0.53	0.54	10000
weighted avg	0.93	0.95	0.94	10000

Table 4-6 confusion matrix, unmatched condition sizes

confusion matrix	predicted 0	predicted 1
true 0	9441	137
true 1	390	32

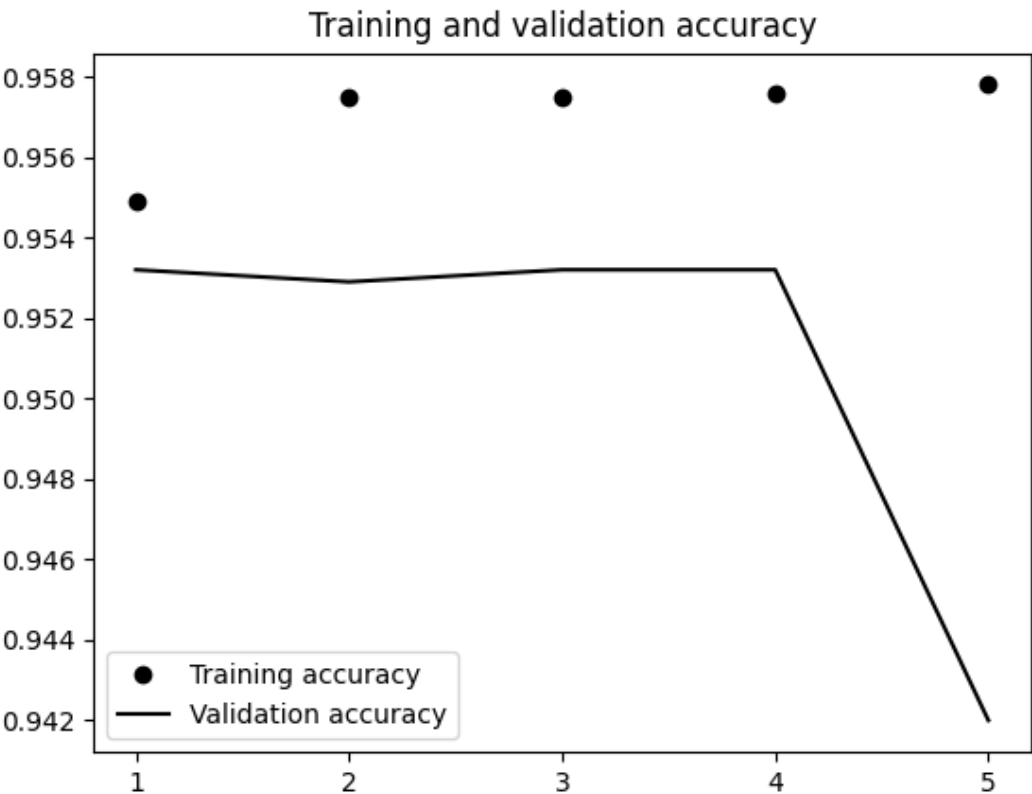
The scatter plot of predictions, below, shows that no posts were predicted as metaphoric with much more than a probability of around 0.7: the algorithm has not learned to predict metaphoric posts very confidently. In comparison the density of points at the bottom left of the plot shows that many non-metaphoric posts were predicted correctly and with high confidence.

Figure 4-2 scatter plot of predictions, unmatched condition sizes



The following plot of training and validation accuracy over five epochs shows that while training accuracy increases in every epoch, validation accuracy drops after the fourth epoch, suggesting that overfitting might be starting to occur. In practice it was found that accuracy was highest at around five epochs.

Figure 4-3 training and validation accuracy



4.3.3.2.2 matched condition sizes

The following table details the precision and recall for the composite *metaphors* condition investigated above, when metaphoric terms are first removed from the text passed to the algorithm, and metaphoric and non-metaphoric condition sizes are matched.

The recall values show that when condition sizes are matched, more metaphoric (81%) than non-metaphoric (70%) posts are predicted correctly. The subsequent scatter plot of the predictions shows that posts are classified over the whole range of scores, from 0 to 1, but with a concentration at the extreme values for each predicted class. There is however also a concentration of points at both ends of the data labelled as 0 (non-metaphoric), which is on the left of the plot, showing that a notable number of posts are confidently predicted as metaphoric when they are in fact labelled as non-metaphoric: metaphoric

posts are being overpredicted. However, performance is far better balanced between the conditions than when unmatched condition sizes were considered in the previous section. For this reason, condition sizes are matched in future analysis, and the analytic focus is solely on the F1 weighted average which as discussed above takes the accuracy of both conditions into account.

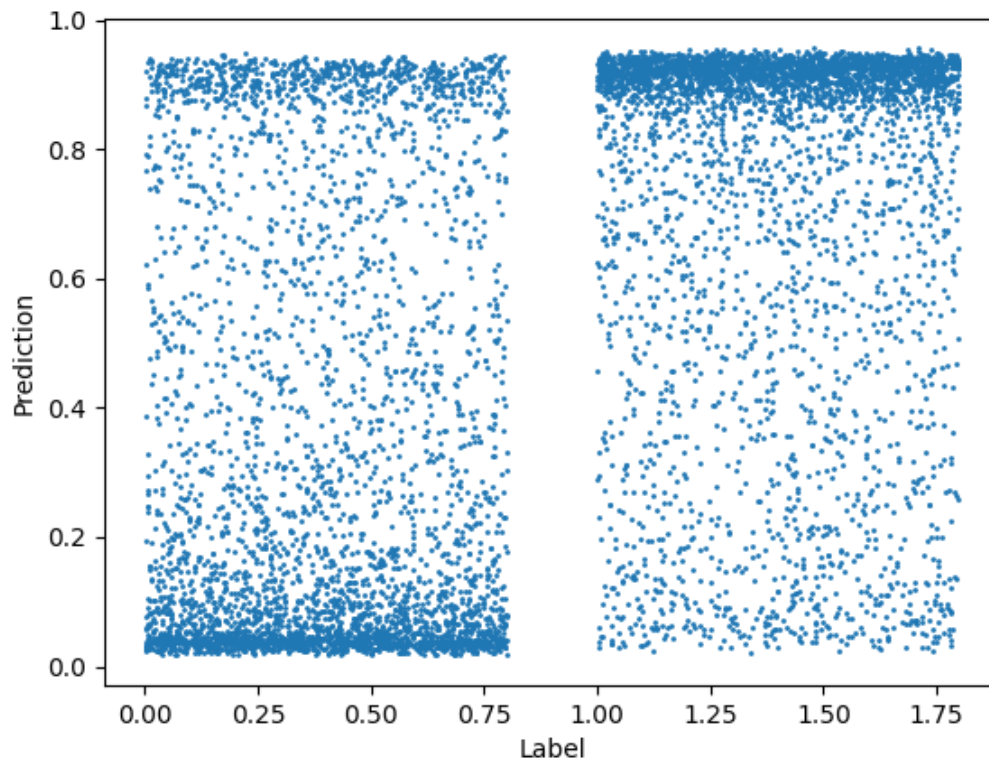
Table 4-7 classification report, matched condition sizes

	precision	recall	f1-score	support
0	0.78	0.70	0.74	3886
1	0.74	0.81	0.77	4052
accuracy			0.75	7938
macro avg	0.76	0.75	0.75	7938
weighted avg	0.76	0.75	0.75	7938

Table 4-8 confusion matrix, matched condition sizes

confusion matrix		predicted 0	predicted 1
true 0		2724	1162
true 1		788	3264

Figure 4-4 scatter plot of predictions, matched condition sizes



The following plot of training and validation accuracy at each epoch for this same scenario (matched condition sizes, and metaphoric terms first removed from the text data passed to the algorithm) shows that while training accuracy continues to improve over the five epochs, validation accuracy falls after the fourth epoch, suggesting that overfitting might be occurring: this supports the use of around five epochs. As with the same analysis for unmatched condition sizes discussed above it was found in practice that use of more than five epochs resulted in worse performance. The subsequent plot from a different iteration of this model over ten epochs supports this view.

Figure 4-5 training and validation accuracy over 5 epochs

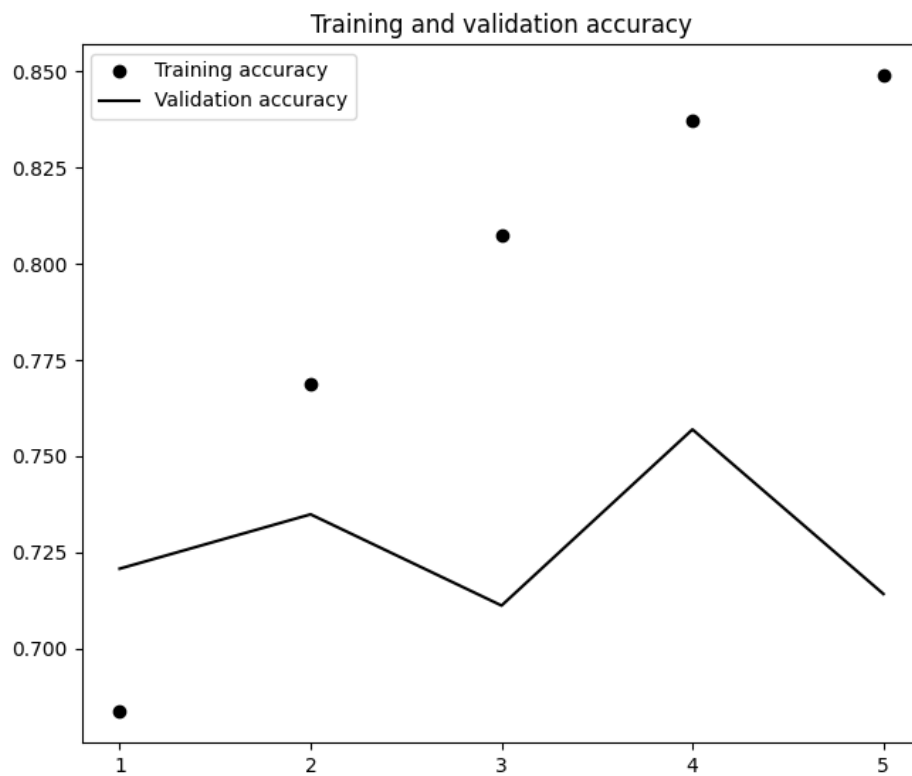
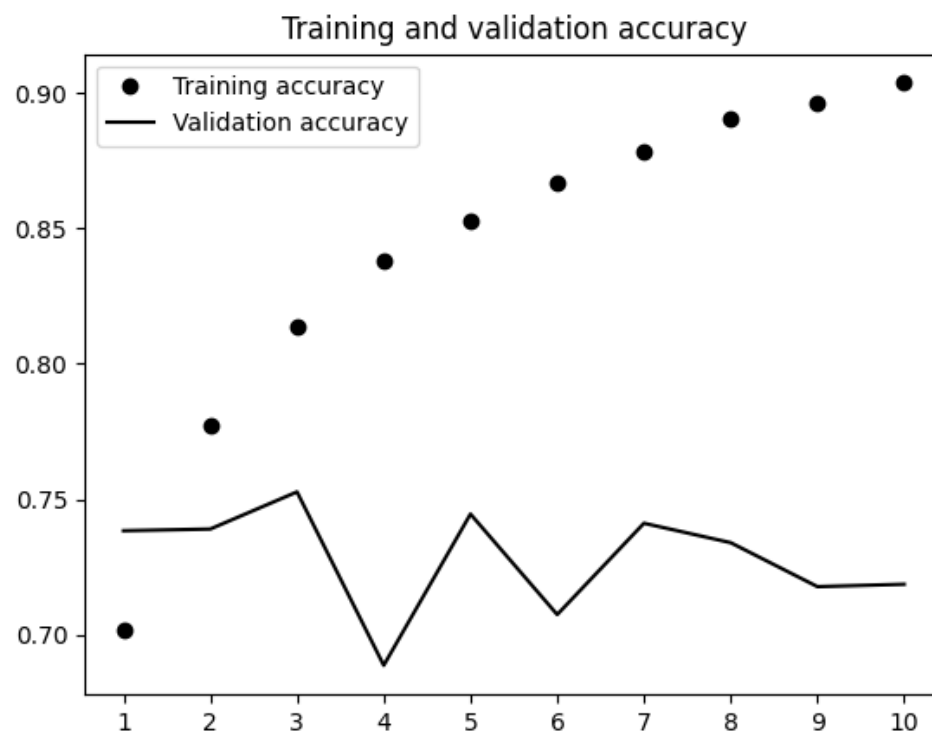


Figure 4-6 training and validation accuracy over 10 epochs



4.3.3.3 Labelling method one (NOUNS), different metaphoric conditions

For the analysis in this section each of the different metaphor themes was explored separately, and in addition, to consider whether it is correct to organise them as separate themes, combinations of the established themes were also considered. If the consideration of multiple themes together results in a higher prediction accuracy, this would indicate that those combined themes are surrounded by a similar language style, and therefore can be considered to do similar social and psychological work. Following on from the results and analysis of the previous section, in this section the model was run for matched condition sizes (number of posts) only, and the metaphoric term(s) were removed from the text data before it was passed to the supervised algorithm.

Combined conditions were labelled by taking the first letter of each metaphor theme included in the analysis. Results in the following table are displayed in descending order of accuracy. In addition, for conditions where the number of test posts in each condition is less than 1,000, because of the consequent potential variability of results due to the distribution of this limited number of posts into training and test sets five iterations of the model were run, with the average of the five weighted average F1 results reported.

Table 4-9 prediction of posts containing particular metaphor theme terms as nouns

metaphoric condition	remove/ retain	match sizes?	weighted average F1	test posts 0	test posts 1
journey	remove	True	0.74	2,403	2,373
rollercoaster	remove	True	0.74	502	502
jr	remove	True	0.74	2,807	2,813
fb	remove	True	0.73	1,370	1,322
fight	remove	True	0.72	648	648
battle	remove	True	0.71	652	684
jf	remove	True	0.71	3,531	3,585

It can be seen from the results table above that when the *journey* and *fight* metaphor themes are considered together (*jf*), although more metaphoric posts are available for

training and testing, the accuracy of the model is decreased (F1=0.71) in comparison to when they are considered separately (F1=0.74 and F1=0.72 respectively). In comparison, when *fight* and *battle* are considered together F1 is increased slightly to 0.73, compared to 0.72 and 0.71 respectively when they are considered separately. When *journey* and *rollercoaster* are considered together, there is no change in the accuracy of the model.

4.3.3.4 Labelling method two: NOUNS_VERBS

In this section the labelling method NOUNS_VERBS was used, which includes metaphoric terms instantiated as both nouns and verbs. The NOUNS_VERBS labelling method was investigated for the *fight* and *battle* metaphor themes only, since, as was shown above, the other themes have minimal or no representation of their constituent terms as verbs. For each theme investigated five iterations of the model were run, with the average of the weighted average F1 results reported.

Table 4-10 labelling methods one (NOUNS) and two (NOUNS_VERBS) compared

	NOUNS	test posts	test	NOUNS_VERBS	test posts	test posts
	weighted	0	posts 1	weighted average	0	1
	average F1			F1		
fight	0.72	648	648	0.70	2,432	2,346
battle	0.71	652	684	0.70	989	973

It can be seen from the results table above that even though the number of posts available for training and testing of the model is increased, the model for both the *fight* and *battle* metaphor themes is slightly less accurate when constituent metaphoric terms are included in verb as well as noun form (on the right of the table) compared to when they are included in noun form only (on the left of the table). This suggests that there are differences in language style surrounding the verb and noun forms of these metaphors, such that they may support different social and psychological work.

4.3.4 Predictions made by the text-based model: all metaphor themes combined

The composite metaphoric condition *metaphors*, explored above, was used for the investigation in this section. Since it has been shown to be more accurate in predicting metaphoric posts, labelling method one (NOUNS) was used in the investigation in this section into the posts that are 'incorrectly' classified by the model: posts that were predicted as metaphoric when labelled as not-metaphoric, and posts that were predicted as not-metaphoric when labelled as metaphoric. Considering why the model makes the 'incorrect' predictions it does gives more insight into the language style surrounding metaphor use.

4.3.4.1 Predictions when the metaphoric noun is retained in the data

As was shown above, when the metaphoric noun is retained in the text passed to the supervised algorithm the accuracy is very high, with a weighted average F1 score of 0.97. In this case, each of the posts classified as 1 (contains the relevant metaphor theme(s)) when labelled as 0 (does not contain the relevant metaphor theme(s)) contains a word from the metaphor dictionary, but not as a noun. For example, in the phrase *rollercoaster ride*, which was labelled as not metaphoric but predicted by the supervised model to be metaphoric, *rollercoaster* is an adjective; in the phrase *my mothers journey*, which was labelled as not metaphoric but predicted by the supervised model to be metaphoric, *journey* is a verb. The retained metaphoric noun, which is the basis for the labelling of posts, provides a lot of information to the algorithm, and this is not helpful in terms of considering what the characteristics of metaphoric posts are aside from the specific metaphoric terms they contain.

4.3.4.2 Predictions when the metaphoric noun is removed from the data

When the metaphoric noun is removed from the text passed to the supervised algorithm, the algorithm must focus more on the wider difference in language style between posts labelled 1 and posts labelled 0. In the analysis described above, when the metaphoric noun was removed from the text, the weighted average F1 score was 0.94 (unmatched condition sizes), which is lower than when the metaphoric noun was retained ($F1=0.97$), but still highly accurate. However, the purpose of this analysis is not just to obtain the highest possible accuracy score, but to investigate what the predictions of the model can tell us about the use of metaphor. The unmatched condition sizes approach is used for this analysis because as discussed above it underpredicts posts as metaphoric: it is anticipated on this basis that those posts predicted metaphoric will provide strong information about the characteristics of metaphoric posts.

In this case where the metaphoric noun was removed from the text passed to the supervised algorithm there is a clear difference in stance between a sample of posts predicted by the supervised learning model to contain the relevant metaphor theme when labelled as not containing it, compared with a sample of posts predicted by the model to not contain the relevant metaphor theme when labelled as containing it, a difference in stance which aligns with the first example of metaphor use in conjunction with language style set out at the start of chapter one. The supervised model, which based its prediction on the language style of any post, predicted that the more discursive and figurative posts with a focus more orientated towards the other would be metaphoric, while it predicted as non-metaphoric those posts that are more literal, and more focused on the current lived experience of their writer. Examples of 'incorrect' predictions made by the model are detailed in the following two sections.

4.3.4.2.1 Posts predicted 1 when labelled 0

A random sample was made of three of the 137 posts predicted to contain at least one of the metaphor themes when labelled as not containing any of the themes, and these are listed in the table below. The first sampled post contains the metaphor *i too am sailing in the same boat*; the second post is an extended fight metaphor which however does not contain any nouns from the metaphor themes; while the third post, described as an apache indian reading, also contains metaphor, for example *each of you will be the shelter for each other*.

Table 4-11 posts predicted as metaphoric when labelled as not metaphoric

post	excerpt from post
1.	how are you coping love. i wish i had some divine insight to give you. but i too am sailing in the same boat .
2.	i choose to accept that the cancer has got there and the stage it at and now i am a cool-headed and strong "commander" in the war zone with my family by me for reinforcement against the big c enemy . i have read up on conventional approaches and explored the so-called lifestyle changes . i have tried to pinpoint what tactics could make a difference and will learn more i have no doubt to try to outwit the enemy . i have a plan i am armed with chemo drugs which is like going nuclear and i have my resilience tactics in place .
3.	this is a non-religious reading from an unknown apache indian - it was the first one i came to and seems lovely. so hopefully you will think so too if not let me know. and i will look some more. apache blessing - author unknown. now you will feel no rain for each of you will be the shelter for each other . now you will feel no cold for each of you will be the warmth for the other in this way you can ride out the storms when clouds hide the face of the sun. in your lives - remembering that even if you lose sight of it for a moment the. sun is still there . and if each of you takes responsibility for the quality of your life together it will be marked by abundance and delight . i wish you both all the best of luck for the future. p.s. and if it is you doing the reading let me know. and i will re-arrange with a few our and us replacing the you. p p s we will want to see a photo on here. much love

4.3.4.2.2 Posts predicted 0 when labelled 1

A sample of the 390 posts predicted to not contain any of the metaphor themes when labelled as containing at least one of the themes, in contrast, although they all contain a noun from the metaphor themes, are all focused on more literal detail, and on the writer's current experience of living with cancer. For example:

its just been a bit hard . i've not had any contact or support throughout this journey so far . i've rung macmillan. but basically i've . muddled through . i think now i perhaps need a bit of help

4.3.5 Section summary: supervised learning text-based method

The text-based supervised learning model had good accuracy in the condition in which the metaphoric term is removed from the data, such that that specific very informative data is not passed to the supervised algorithm. Accuracy decreased to 0.71 when the *journey* and *fight* themes were considered together in the *jj* condition, from 0.74 and 0.72 respectively, supporting the insight that these metaphor themes are surrounded by different language styles, to perform different work.

When metaphoric terms were included as verbs as well as nouns the accuracy of the model in predicting the metaphoric conditions *fight* and *battle* decreased to 0.70 from 0.72 and 0.71 respectively, even though more posts were available for training. This suggests that metaphoric nouns and verbs have a different surrounding language style: they are potentially doing different social and psychological work. It may also indicate, as discussed in chapter two, that metaphor instantiated as nouns is more impactful.

A sample of posts predicted to contain metaphor when labelled as not containing metaphor were all found to contain other metaphor that was not represented in the established metaphor themes. In addition, those posts were found to be more discursive and figurative and to address support towards the other. A sample of posts predicted to not contain metaphor when labelled as containing metaphor in comparison was found to have an overall more literal focus even though they each contained a metaphoric term from the themes instantiated as a noun. The supervised model located this difference in style. These findings support the thesis and example set out at the start of chapter one,

that use of metaphor to characterise a concept co-occurs with a language style that has been associated with better mental health, which may be more directed towards the other, while posts that do not use such metaphor have a language style that has been associated with worse mental health, such that the focus is on the current literal lived experience of the participant.

In the next section the relationship between language style and metaphor use is explored in a more finely grained analysis which, instead of supplying whole post texts to the supervised algorithm, instead supplies the counts of a wide range of linguistic variables to the algorithm. Again, individual metaphor themes, and groupings of metaphor themes, are considered.

4.4 Count-based method (COUNT)

In the previous section, the text from each post was passed as input data to the supervised algorithm. In this section in comparison, for each post 143 separate lexical items are counted. It is those counts that are passed to the supervised algorithm, and not the actual text of the post. Although the presence of each metaphor theme is also counted for each post, this is used to label the data before it is passed to the supervised algorithm - the counts of the metaphor themes are not passed to the algorithm.

4.4.1 Counting lexical items in each post

For each post, the 143 linguistic variables described in the table below are counted. The schemas from which these variables are sourced are described more fully in subsequent sections. A correlation analysis of linguistic variables with each other allowed variables that correlate exactly with other variables to be discarded. Some of the variables in the

following table appear to be repeats, for example CCONJ and CC, however, they have been found to do slightly different work in practice.

Table 4-12 linguistic variables counted for analysis

	linguistic variable	description
1.	Positive	words with a positive emotional valence
2.	Negative	words with a negative emotional valence
3.	Anger	Anger words
4.	Anticipation	Anticipation words
5.	Disgust	Disgust words
6.	Fear	Fear words
7.	Joy	Joy words
8.	Sadness	Sadness words
9.	Surprise	Surprise words
10.	Trust	Trust words
11.	ADJ	adjective
12.	ADP	adposition
13.	ADV	adverb
14.	AUX	auxiliary verb
15.	CCONJ	coordinating conjunction
16.	DET	determiner
17.	INTJ	interjection
18.	NOUN	noun
19.	NUM	numeral
20.	PART	particle
21.	PRON	pronoun
22.	PROPN	proper noun
23.	SCONJ	subordinating conjunction
24.	SYM	symbol
25.	VERB	verb
26.	X	other
27.	punctuation	punctuation
28.	\$	symbol, currency
29.	ADD	email
30.	AFX	affix
31.	CC	coordinating conjunction
32.	CD	cardinal number
33.	DT	determiner
34.	EX	existential there
35.	FW	foreign word
36.	IN	conjunction, subordinating or preposition

37.	JJ	adjective
38.	JJR	adjective, comparative
39.	JJS	adjective, superlative
40.	LS	list item marker
41.	MD	verb, modal auxiliary
42.	NN	noun, singular or mass
43.	NNP	noun, proper singular
44.	NNPS	noun, proper plural
45.	NNS	noun, plural
46.	PDT	predeterminer
47.	POS	possessive ending
48.	PRP	pronoun, personal
49.	PRP\$	pronoun, possessive
50.	RB	adverb
51.	RBR	adverb, comparative
52.	RBS	adverb, superlative
53.	RP	adverb, particle
54.	TO	infinitival 'to'
55.	UH	interjection
56.	VB	verb, base form
57.	VBD	verb, past tense
58.	VBG	verb, gerund or present participle
59.	VCN	verb, past participle
60.	VBP	verb, non-3rd person singular present
61.	VBZ	verb, 3rd person singular present
62.	WDT	wh-determiner
63.	WP	wh-pronoun, personal
64.	WP\$	wh-pronoun, possessive
65.	WRB	wh-adverb
66.	XX	unknown
67.	acl	clausal modifier of noun (adjectival clause)
68.	acompl	adjectival complement
69.	advcl	adverbial clause modifier
70.	advmod	adverbial modifier
71.	agent	agent
72.	amod	adjectival modifier
73.	appos	appositional modifier
74.	attr	attribute
75.	aux	auxiliary
76.	auxpass	auxiliary (passive)
77.	case	case marking

78.	cc	coordinating conjunction
79.	ccomp	clausal complement
80.	compound	compound
81.	conj	conjunct
82.	csubj	clausal subject
83.	csubjpass	clausal subject (passive)
84.	dative	dative
85.	dep	unclassified dependent
86.	det	determiner
87.	doobj	direct object
88.	expl	expletive
89.	intj	interjection
90.	mark	marker
91.	meta	meta modifier
92.	neg	negation modifier
93.	nsubj	nominal subject
94.	nsubjpass	nominal subject (passive)
95.	nummod	numeric modifier
96.	oprd	object predicate
97.	parataxis	parataxis
98.	pcomp	prepositional complement
99.	pobj	object of a preposition
100.	poss	possession modifier
101.	preconj	preconjunct
102.	prep	prepositional modifier
103.	prt	phrasal verb particle
104.	quantmod	quantifier phrase modifier
105.	relcl	relative clause modifier
106.	xcomp	open clausal component
107.	i	
108.	me	
109.	you	
110.	he	
111.	him	
112.	she	
113.	her	
114.	it	
115.	we	
116.	us	
117.	they	
118.	them	

119.	myself	
120.	yourself	
121.	himself	
122.	herself	
123.	itself	
124.	themselves	
125.	ourselves	
126.	yourselves	
127.	themselves	
128.	mine	
129.	his	
130.	hers	
131.	its	
132.	ours	
133.	yours	
134.	their	
135.	my	
136.	your	
137.	our	
138.	theirs	
139.	fpp	first-person plural composite variable, includes we, us, our, ours, and ourselves
140.	unique_words	unique words
141.	long_words	words > 7 characters in length
142.	monosyllabic_words	monosyllabic words, words 1 syllable in length
143.	polysyllabic_words	polysyllabic words, words > 2 syllables in length

4.4.1.1 Universal Part-of-speech tags

spaCy maps all language-specific part of speech tags to a small, fixed set of word type tags following the Universal Dependencies scheme. The universal tags don't code for any morphological features and only cover the word type. The universal part-of-speech tags listed below are used in this analysis. The SPACE tag is excluded in the current investigation since it provides information about the number of words in a post, which is too much information to pass to the supervised algorithm: it equates to word count, which is typically much higher in metaphoric posts in comparison to non-metaphoric posts.

ADJ, ADP, ADV, AUX, CONJ, CCONJ, DET, INTJ, NOUN, NUM, PART, PRON, PROP, PUNCT, SCONJ, SYM, VERB, X, SPACE

4.4.1.2 English Part-of-speech tags

The spaCy English part-of-speech tagger uses the OntoNotes 5 version of the Penn Treebank tag set. The `_SP` tag, which represents a space, is not used in the analysis.

\$, ADD, AFX, CC, CD, DT, EX, FW, GW, HYPH, IN, JJ, JJR, JJS, LS, MD, NFP, NIL, NN, NNP, NNPS, NNS, PDT, POS, PRP, PRP\$, RB, RBR, RBS, RP, SP, SYM, TO, UH, VB, VBD, VBG, VBN, VBP, VBZ, WDT, WP, WP\$, WRB, XX, `_SP`

4.4.1.3 Syntactic Dependency Parsing (English)

The following spaCy tags relate to syntactic dependency parsing.

acl, acomp, advcl, advmod, agent, amod, appos, attr, aux, auxpass, case, cc, ccomp, compound, conj, cop, csubj, csubjpass, dative, dep, det, dobj, expl, intj, mark, meta, neg, nn, nounmod, npmod, nsubj, nsubjpass, nummod, oprd, obj, obl, parataxis, pcomp, pobj, poss, preconj, prep, prt, punct, quantmod, relcl, root, xcomp

4.4.1.4 Pronouns

Because of established findings regarding pronoun use and state of mind discussed in chapter one, in the current investigation the spaCy generic `-PRON-` tag is split into the following variables, which are counted separately:

i, me, you, he, him, she, her, it, we, us, they, them, myself, yourself, himself, herself, itself, themselves, ourselves, yourselves, themselves, mine, his, hers, its, ours, yours, theirs, my, your, his, her, its, our, their, `*fpp`

`*fpp` is an additional grouped variable that includes the following first-person plural pronouns:

we, us, our, ours, ourselves

4.4.1.5 Emotions and positive/negative valence

The NRC (Canadian National Research Council) emotion lexicon (Mohammad, 2021), a crowd-sourced resource of words categorised as having positive or negative valence, and belonging in eight different emotion categories (anger, anticipation, disgust, fear, joy,

sadness, surprise, and trust), is used here to identify and count emotion words. The NRC emotion lexicon was used in a quantitative investigation into metaphor and emotion, which found that metaphoric language tends to have a stronger emotional impact than does literal language (Mohammad, Shutova and Turney, 2016).

The lemmas representing each metaphor theme are also present in various emotion categories, for example, *fight* is in the *Negative*, *Anger*, and *Fear*, categories. In order to be able to consider the meaning of emotion words in various metaphoric conditions, the lemmas from the metaphor theme under investigation are first removed from the text before counts of the linguistic variables are made. So, for example, for the posts in which the *journey* theme is present, the *journey* lemmas are removed from that post text before linguistic variables are counted, but the *fight* lemmas are not removed from that post text. In addition, the term *cancer*, since it is a dominant word with specific meaning in this context, was removed from all emotion category lists.

4.4.1.6 Other variables

The spaCy-related package textacy is used here to count, for each post, the numbers of unique words, long words, monosyllabic words, and polysyllabic words.

4.4.2 Word count: variable counts converted to proportions

The average length of a metaphoric post when all the metaphor themes are considered together is 217 tokens. In comparison the average length of a non-metaphoric post is 108 tokens. This is a lot of information to provide to the algorithm and may lessen the focus on other linguistic variables. To account for this, as discussed above, in the current investigation the word count of a post is excluded from the data passed to the supervised

algorithm. In addition, all the other linguistic variables for any post are expressed as a proportion of the total word count of that post.

4.4.3 Selection of supervised machine learning models

In the previous section, in which post texts were passed to the supervised algorithm, the Keras deep learning Python library was used to set up a deep learning model. In this section, the scikit-learn machine learning Python library is used, since it was found to produce better results with the numeric data used here.

In order to find the best algorithm to use with this data, the data was supplied to eight typical classification models from the scikit-learn library. The default parameters were used for each model, except in the case where the linear kernel was used for the support vector machine classifier, for which RBF is the default kernel.

The following table ranks the models by their accuracy in predicting the metaphoricity of posts from the cancer.net data.

Table 4-13 classification models ranked by accuracy

F1 score	classification model
0.748566	Support Vector Machine, RBF kernel
0.744896	Multi-Layer Perceptron
0.740766	AdaBoost
0.735949	Support Vector Machine, Linear kernel
0.731590	Gaussian Process
0.727460	Random Forest
0.667125	Decision Tree
0.629961	Naive Bayes
0.610920	Nearest Neighbours

4.4.3.1 Grid search with cross validation of best models

For the two highest performing models found above, Support Vector Machine with RBF kernel, and Multi-Layer Perceptron, a grid search was carried out, in which a range of

parameters was tested to find the combination of parameters providing the highest weighted average F1 score. The highest performing model was the Multi-Layer Perceptron, and this model is used in subsequent analysis in this section, with the following parameters:

```
{'activation': 'tanh', 'alpha': 0.05, 'hidden_layer_sizes': (100,), 'learning_rate': 'adaptive', 'max_iter': 1000, 'solver': 'adam'}
```

4.4.4 Stratified k-fold

Because the data supplied to the supervised model is shuffled before being split into training, validation, and test data, results may be different for different iterations of the model. Disparity between iterations is less of an issue where many data records are available, as is the case here for the overarching *metaphors* condition. It becomes more problematic when the *rollercoaster* metaphor theme is investigated since there are relatively few posts counted as containing this theme. This issue is addressed here by use of *StratifiedKFold* from the scikit-learn library. This divides the data into k groups of samples, called folds, of equal sizes (if possible). The prediction function is learned using folds, and the fold left out is used for testing the model, such that each fold is used once for testing. The F1 score reported is the average of the weighted average F1 scores for each fold in the sample.

4.4.5 Count-based method, analysis by labelling type and metaphor theme

The following analyses using the count-based supervised learning method all use matched condition sizes so that the weighted average F1 result is easier to interpret and compare between conditions; in this case, as was shown in the previous section, the weighted average F1 is the same as the macro weighted average. When the conditions are balanced in size, any F1 result > 0.5 shows that the supervised learning algorithm is doing some work to improve on a random prediction of results. For example, in a trial in which

a single variable supplied to the algorithm was filled with random integers, the weighted average F1 score was always below 0.50, and on 10 trials was always within the range 0.33 to 0.43, had a wide standard deviation, and did not vary between metaphor conditions.

The following table lists in order of weighted average F1 score the results when labelling method one (NOUNS) was used to label the different metaphoric conditions. Results for labelling method two (NOUNS_VERBS) are also included in the table for comparison for the *fight* and *battle* conditions, since they are the metaphoric conditions that also have a substantial instantiation as verbs. With the linguistic variable counting method currently being investigated there was found to be no notable difference in score between remove and retain conditions, so the remove method is used here, as it was with the TEXT method in the previous section, i.e. the metaphoric noun was removed from the text before counts of the 143 linguistic variables were made.

Table 4-14 weighted average F1 accuracy of model in different metaphoric conditions

metaphoric condition	NOUNS weighted average F1	precision	recall	test posts per condition (matched)	NOUNS_VERBS weighted average F1	test posts per condition (matched)
journey	0.81	0.79	0.84	1,433		
metaphors	0.80	0.78	0.81	2,381		
jr	0.80			1,686		
jf	0.79			2,135		
fight	0.77	0.73	0.82	389	0.77	1,434
fb	0.77			807	0.77	2,023
roller coaster	0.73	0.72	0.78	301		
battle	0.71	0.70	0.73	401	0.76	589

It can be seen from the results table above that the model was most accurate for the *journey* condition (F1=0.81), and least accurate for the *battle* condition (F1=0.71). The model was less accurate when *journey* was combined with other metaphor themes,

supporting the view that *journey* is surrounded by different language, and does different social and psychological work, to those other themes. For all analysis, as for the analysis in the previous text-based method in which condition sizes were matched, precision and recall are fairly well balanced, and recall is consistently higher than precision. This supports the continued use of the higher level weighted average F1 metric to assess the performance of the model.

In comparison to the text-based method explored above (TEXT), with the count-based method (COUNT) under investigation in the current section *rollercoaster* was less accurate when used alone than when combined with *journey* in the *jr* theme; this is most likely due to *journey*, for which there are many more posts, carrying the analysis when the two metaphoric conditions are combined. *Fight* had the same accuracy when used alone as it did when combined with *battle* ($F1=0.77$), but *battle* was less accurate when used alone ($F1=0.71$). However, when *battle* was included as a verb as well as a noun, the accuracy of the model increased to 0.76, which may partly be explained by the increased number of posts available on which to train the model. This made no difference for *fight*, for which the accuracy did not change between the NOUNS and NOUNS_VERBS conditions.

4.4.6 Comparison of results when only pronoun counts are used

As was discussed in chapter one, the use and range of pronouns in a post is particularly pertinent to language style as it relates to state of mind. To explore the predictive power of pronouns for the different metaphoric conditions, the supervised learning model was run for each metaphor condition, but instead of using the counts of all 143 linguistic variables, only the 35 counts relating to pronouns were included. This is contrasted in the results table below with the condition in which all 143 linguistic variable counts were

included (centre column), and the condition in which all but pronoun variable counts were included (righthand column).

Table 4-15 weighted average F1 when only pronoun variables are included compared with other conditions

metaphoric condition	weighted average F1: pronouns only	weighted average F1: all variables	weighted average F1: all except pronouns
journey	0.66	0.81	0.80
fight	0.66	0.77	0.75
battle	0.62	0.71	0.70
rollercoaster	0.60	0.73	0.75

It can be seen from the results table above that the supervised model still had better than average accuracy in predicting the metaphoric condition of a post even when only the counts of the pronouns present in that post were included, each of which was expressed as a proportion of all the words in the post. In addition, the accuracy of the model decreased in the condition in which all but pronoun counts were used, in comparison to the condition in which all linguistic variable counts were used, except for the *rollercoaster* condition: predictive accuracy was better for *rollercoaster* when pronoun counts were not supplied to the algorithm. These results suggest that there is a strong relationship between pronoun use and use of the metaphor themes. They also suggest that the *rollercoaster* theme has a different relation to pronouns than do the other themes.

4.4.7 Comparison with other automatic metaphor identification systems

Although it is now an active area of research in natural language processing, with shared global tasks for comparative evaluation of techniques, automatic metaphor identification still currently consists of a diverse range of techniques that may be drawn on for specific application needs: there is currently no natural language processing standard for identifying metaphor (Veale, Shutova and Klebanov, 2016, p. 121).

At present, no computational approach to blending seems entirely satisfactory, and the field awaits a robust hybrid mode that combines the scale and practicality of

bottom-up statistical approaches with the top-down structured control of symbolic approaches. (Veale, Shutova and Klebanov, 2016, p. 53)

A survey of computational metaphor processing systems undertaken with the aim of defining an evaluation framework for this important work found that there are four main approaches to metaphor detection: linguistic, statistical, classification, and conceptual (Shutova, 2015), all of which have been used in this or earlier chapters in this investigation to provide specific insights relating to metaphor use in the cancer.net community, and to support the supervised classification analysis of the current chapter.

In a summary of the Association for Computational Linguistics 2020 Workshop on Processing Figurative Language metaphor detection task it is noted that deep learning methods such as those explored in the current chapter have become more popular, with more than half of entries for the 2020 workshop based on the transformer language model BERT (Devlin *et al.*, 2019). Introduced in 2017, the transformer is a deep learning model that, in comparison to previous deep learning models commonly used for text such as the LSTM (Long Short Term Memory) model used above in the current investigation, uses only an attention mechanism to access all previous states and weight them according to a learned measure of relevancy to the current token. This provides clearer information about distant relevant tokens than is possible in LSTM models, for example. Since the order of sequential data is ignored in transformer models, much more parallel training can take place, with concomitant reduced training times, and much larger training datasets. This has led to the development of large pre-trained systems such as BERT (Bidirectional Encoder Representations from Transformers) (Devlin *et al.*, 2019).

Although some exploration of transformer models was made in the current investigation, the resources were not available to this project to run these very large models with

sufficient data. A limited exploration using the ELMo transformer model, which is smaller than BERT, with a very small number of training records (500) produced good results, with an F1 accuracy >0.8 . Although this is an area of interest for future investigation, it is not pursued further here for pragmatic reasons, since excellent results are in any case obtained: results that are good enough for the investigative purpose defined here, to assess the technique of considering language style to predict when potentially deliberate noun-based metaphor is and is not used in a post, and to determine which metaphoric terms belong together in a single theme. The results obtained here also support the focus on metaphor instantiated as nouns, although for the COUNT method the *battle* theme was predicted more accurately when metaphors instantiated as both nouns and verbs were included.

The accuracy of the results obtained in the current study from the very diverse, naturally occurring data of cancer.net is very high when compared with the methods discussed in the Shutova survey (Shutova, 2015), for which available accuracy results are collected in the table below. The F1 results for the very specific tasks of the Association for Computational Linguistics 2020 Workshop on Processing Figurative Language metaphor detection similarly cover a very wide range (Leong *et al.*, 2020). In conclusion, the labelling method used in the current study is valid for splitting data based on the presence or absence of metaphor themes in each post, for the purpose of further analysis. The method explored in the current chapter may also be productive for locating metaphor in a large text corpus, which as discussed previously is an important issue for metaphor research (Gibbs, 2017).

Table 4-16 automatic metaphor identification F1 results in order of accuracy

author	F1 score	description
current investigation, count method	0.81	cancer.net, journey theme
current investigation, count method	0.80	cancer.net, all metaphor themes combined
(Shutova and Sun, 2013)	0.79	
(Tsvetkov, Mukomel and Gershman, 2013)	0.78 (0.72-0.85)	
(Hovy et al., 2013)	0.75	
(Li and Sporleder, 2009)	0.75	
(Mohler et al., 2013)	0.70	
(Li, Zhu and Wang, 2013)	0.69	is-a metaphor
	0.58	verbal metaphor
(Turney et al., 2011)	0.68	
(Wilks et al., 2013)	0.67	
(Heintz et al., 2013)	0.59	
(Krishnakumaran and Zhu, 2007)	0.58	
(Birke & Sarkar, 2006)	0.54	

4.4.8 Section summary: supervised learning

Machine learning techniques were found to be accurate in finding the dominant metaphor themes in the diverse cancer.net text corpus. Unsupervised learning techniques used in previous chapters exposed previously unknown metaphoric themes and language.

Supervised learning in the current chapter built on that new knowledge about metaphoric language in the data, producing a model that was very accurate in predicting the condition of a set of unlabelled test posts, and which provided more insights into metaphor use in this data. In particular it confirmed that posts containing the metaphor themes are likely to be more figurative and discursive, and to be addressed towards the other, while posts that do not contain the metaphor themes are likely to be more literal and focused on the current lived experience of their speaker.

Use of supervised learning to predict the presence of different metaphoric conditions in a post has shown that each metaphoric condition is predictable by the trained supervised

model based on the language style of the posts in which it is used. It was also demonstrated that the *journey* theme is the most recognisable by the model and is more recognisable when considered alone than when considered in conjunction with other metaphor themes. This suggests that there is a particular distinctive language style surrounding use of the *journey* theme. It is also notable that *rollercoaster* is the only metaphoric condition in which predictive accuracy was decreased when counts of pronouns were included in the training data supplied to the supervised algorithm.

The analysis in which different combinations of metaphor themes were considered suggests that the different metaphor themes do different work, which is reflected in the different language styles that typically surround them. The following organisation of metaphor themes has therefore been validated.

Table 4-17 metaphor themes and constituent nouns

metaphor theme	nouns used to invoke the theme
journey	journey
rollercoaster	rollercoaster, coaster
fight	fight, fighter
battle	battle, battlefield

The results summary table below shows that the COUNT method was notably more accurate than the TEXT method at predicting the *journey* and *fight* themes; the TEXT method was slightly more accurate at predicting the *rollercoaster* theme; and there was no difference between the two methods for predicting the *battle* theme. This also suggests a difference in the language style surrounding the different themes, and that this may require different methods for optimal detection in unlabelled data. It also supports the use of multiple methods for the current analysis. Where the accuracy of the count-based method is higher this suggests that it is the proportions of different linguistic variables that more accurately represent the work of the metaphor theme under consideration.

Where the accuracy of the text-based method is higher, this suggests that a focus on the

positioning of terms within a text, and the relationships between terms, including semantic similarity, more accurately represents the work of the metaphor theme under consideration.

Table 4-18 weighted average F1 accuracy of model in different metaphoric conditions, TEXT and COUNT methods

metaphoric condition	TEXT method weighted average F1	COUNT method weighted average F1	difference
journey	0.74	0.81	0.7
fight	0.72	0.77	0.5
roller coaster	0.74	0.73	0.1
battle	0.71	0.71	0

In the next section, the NOUNS labelling method explored and validated in the current section is used to split the cancer.net data into two corpora based on the presence or absence of community metaphor themes. The aim is to understand in more detail the differences in language style between posts where particular metaphor themes are and are not present, and to assess the statistical significance of any such differences. Use of unsupervised and supervised machine learning techniques to investigate metaphor in cancer.net is now complete.

4.5 Comparing language style between metaphoric conditions

For the analysis in this section the cancer.net data is split into two based on the presence or absence of each metaphor theme in any single post, with corpus F, the corpus of focus, consisting of all cancer.net posts that contain at least one instance of the metaphor theme under investigation, and corpus R, the reference corpus, consisting of all cancer.net posts that do not contain the metaphor theme under investigation. The frequency per 1 million words of each of 143 linguistic variables in each corpus is counted, and the significance of the difference between these is calculated. There are many significant results at the $p < 0.05$ level, so variables are only listed in the output if the log likelihood (LL in the

tables) of the difference between the corpora for any linguistic variable is > 15.13 (significant at $p < 0.0001$). Log likelihood was selected for this purpose because of its standard use in the area of linguistics to test the significance of differences between corpora, and for the current application, in comparison to the consideration of keywords undertaken in chapter two, there is not the issue that log likelihood may identify false positives (Brezina, 2018; Kilgarriff, 2009). The score is the simple maths parameter, which is included for consistency with the keyword analysis of chapter two. The simple maths parameter is calculated here on the same basis described for the Sketch Engine keywords and key terms analysis in chapter two, with the smoothing parameter N set to 1.

To constrain the output further, those variables are only listed if the effect size Cohen's d is also > 0.4 , where Cohen's d values greater than 2 and less than 5 represent a small effect, and values greater than 5 represent a medium effect. In conjunction with p values, effect size is an important consideration since it takes into account sample size, which is of particular relevance when sample sizes are large. Results are ordered by proportion of change between the two corpora, with the direction of change in corpus F indicated with + or -. Since the focus in the current investigation is on personal pronouns and emotion words as a particular site of psychological work, where such terms are increased in corpus F they are marked in bold; where they are decreased in corpus F they are underlined. However, since this is a linguistics investigation, all significantly different linguistic variables are listed.

In order to consider more specifically the work of a particular metaphor theme used in isolation from other metaphor themes, i.e. when it is not supporting an ideological dilemma, the same analysis described above is also carried out in a comparison in which all metaphor themes other than the specific theme under consideration are excluded. For

this analysis posts are included that contain only the metaphor theme of focus, or that contain none of the metaphor themes. To support this analysis the following table details the presence and overlap of each theme. Exclusive posts are those posts that contain that specific metaphor theme, but no other metaphor theme, while overlap posts refers to the number of posts containing that specific theme which also contain at least one other metaphor theme. It can be seen from the table that each metaphor theme is typically used alone, and this is most true of the *journey* theme, which is used in isolation from other metaphor themes in 93% of instances.

Table 4-19 distribution and overlap of metaphor themes

metaphor theme	posts	% of all posts	exclusive posts	% exclusive posts	overlap posts
journey	7,164	2.75	6,643	93	521
rollercoaster	1,507	0.58	1,233	82	274
fight	1,945	0.75	1,660	85	285
battle	2,004	0.77	1,693	84	311
any theme	11,906	4.57			

In addition, since established research findings discussed in chapter one relate diversity of pronoun use, and relatively increased use of positive emotion words, to better mental health, for each comparative corpus the average standard deviation of pronouns in a post is calculated, as well as the ratio of positive to negative emotion words. Counts of emotion variables do not include terms associated with the particular metaphor theme under investigation, for example *journey*, or *fight*.

4.5.1 journey metaphor theme

For the analysis in this section the data was split based on the presence or absence of the *journey* metaphor theme. Corpus F consists of all the posts that have been counted as containing at least one instance of the *journey* theme, while corpus R consists of all the posts that do not contain at least one instance of the *journey* theme. The first results table

below shows that a wider range of pronouns is used in posts in which the *journey* theme is present, and there is also a higher ratio of positive to negative emotion words.

Table 4-20 density of journey theme in cancer.net

	posts	tokens	average post length (tokens)	standard deviation of pronouns	ratio positive to negative emotion words
Corpus F journey>0	7,164 (2.75%)	1,469,893	205	2.62	1.49
Corpus R journey=0	253,077	27,028,314	107	1.63	1.24

The following table lists the 61 variables that were found to be significantly different in density between the two corpora, ranked according to the proportion of change between corpus F and corpus R. It can be seen that the first-person plural pronouns *our*, *us*, *fpp*, and *we*, are the most increased in corpus F in comparison to corpus R, while the first-person singular pronoun *i*, and *Negative* emotion words, are the most decreased key variables. This is a language style that has been associated with better mental health. In addition, the positive emotion variables *Joy*, *Trust*, and *Anticipation*, are increased, while the negative emotion variables *Sadness* and *Fear* are decreased.

It can be seen that both monosyllabic and polysyllabic words are significantly increased in posts in which *journey* is used compared to posts in which it is not used. This is possible because the textacy Python library used for these calculations by default sets 3 as the number of syllables a word must have to be counted as polysyllabic, and this is the value used in the current investigation: the monosyllabic and polysyllabic categories do not include all words analysed.

Table 4-21 most changed linguistic variables between posts that do and do not contain journey

	variable	freq F	freq R		% change	LL	Cohen's d	score
1.	our	1926	1050	+	83.37	804	0.41	1.83
2.	us	2248	1435	+	56.74	543	0.43	1.57
3.	fpp	10813	6979	+	54.94	2489	0.54	1.55
4.	we	6498	4432	+	46.64	1164	0.47	1.47
5.	your	9147	7286	+	25.55	610	0.62	1.26
6.	you	21204	17823	+	18.97	840	0.66	1.19
7.	JJS	2990	2589	+	15.49	82	0.44	1.15
8.	relcl	12782	11476	+	11.38	199	0.58	1.11
9.	<u>i</u>	34781	38897	-	10.58	631	0.42	0.89
10.	<u>Negative</u>	32507	36318	-	10.49	579	0.44	0.9
11.	WRB	8075	7311	+	10.45	107	0.52	1.1
12.	PRP\$	30725	27865	+	10.26	395	0.66	1.1
13.	Joy	26497	24245	+	9.29	282	0.67	1.09
14.	neg	12853	14145	-	9.14	170	0.42	0.91
15.	CCONJ	44922	41271	+	8.85	437	0.67	1.09
16.	unique words	581416	637645	-	8.82	7136	0.81	0.91
17.	<u>Sadness</u>	20925	22889	-	8.58	242	0.44	0.91
18.	conj	46992	43367	+	8.36	411	0.66	1.08
19.	pcomp	7009	6524	+	7.44	49	0.48	1.07
20.	Positive	48390	45096	+	7.30	327	0.67	1.07
21.	VBN	21362	23026	-	7.23	172	0.46	0.93
22.	poss	30172	28197	+	7.00	188	0.62	1.07
23.	prep	84001	78525	+	6.97	520	0.68	1.07
24.	punctuation	96449	90242	+	6.88	581	0.64	1.07
25.	ADP	99309	92990	+	6.80	585	0.68	1.07
26.	IN	108028	101242	+	6.70	620	0.67	1.07
27.	pobj	76169	71409	+	6.67	432	0.67	1.07
28.	TO	22977	21605	+	6.35	119	0.61	1.06
29.	<u>Fear</u>	24302	25924	-	6.26	145	0.47	0.94
30.	DET	82284	77539	+	6.12	396	0.65	1.06
31.	DT	75808	71451	+	6.10	363	0.65	1.06
32.	long_words	134182	126489	+	6.08	639	0.67	1.06
33.	VB	60816	57549	+	5.68	254	0.69	1.06
34.	WP	5259	4990	+	5.38	20	0.44	1.05
35.	mark	20909	19921	+	4.96	67	0.55	1.05
36.	Trust	32496	31020	+	4.76	96	0.64	1.05
37.	advcl	24424	23336	+	4.66	70	0.60	1.05
38.	VBZ	27843	29141	-	4.45	82	0.55	0.96

39.	attr	12423	11906	+	4.34	31	0.57	1.04
40.	MD	19469	18721	+	4.00	41	0.59	1.04
41.	SCONJ	16397	15769	+	3.98	34	0.50	1.04
42.	VBP	45309	47066	-	3.73	93	0.65	0.96
43.	Anticipation	34709	33527	+	3.53	57	0.67	1.04
44.	ccomp	36892	38186	-	3.39	62	0.57	0.97
45.	it	15496	16004	-	3.17	23	0.49	0.97
46.	AUX	68651	70791	-	3.02	91	0.62	0.97
47.	det	63548	61691	+	3.01	77	0.60	1.03
48.	polysyllabic_words	56804	55171	+	2.96	67	0.58	1.03
49.	nsubj	113623	116904	-	2.81	130	0.61	0.97
50.	NN	134310	138186	-	2.80	153	0.59	0.97
51.	VBG	23360	24029	-	2.79	26	0.54	0.97
52.	ADV	80301	78297	+	2.56	71	0.65	1.03
53.	xcomp	18697	18231	+	2.55	16	0.58	1.03
54.	ADJ	68691	67089	+	2.39	53	0.67	1.02
55.	JJ	62558	61210	+	2.20	41	0.66	1.02
56.	PRON	149515	146330	+	2.18	96	0.66	1.02
57.	advmod	78646	77073	+	2.04	44	0.65	1.02
58.	NOUN	166802	169988	-	1.87	84	0.60	0.98
59.	aux	66525	67783	-	1.86	33	0.61	0.98
60.	amod	38470	37803	+	1.77	16	0.58	1.02
61.	monosyllabic_words	804058	792802	+	1.42	222	0.68	1.01

4.5.1.1 journey theme used in isolation

In the following analysis, to consider further the specific work of the *journey* theme, posts were excluded in which *journey* is used alongside other themes, such that corpus F consists of posts that contain only the *journey* metaphor theme, while corpus R consists of posts that contain none of the metaphor themes. In this case, there are 58 significantly different variables. It can be seen that the most notable difference when all posts containing other metaphor were excluded is that the first-person plural pronoun *our* is no longer significantly different between the two corpora, indicating that *our* is associated with *journey* only when it is used alongside other metaphor themes: it is not significant in the work of the *journey* metaphor used in isolation, but may be related to its use in an ideological dilemma, perhaps in conjunction with a negotiation of subject positions.

Table 4-22 most changed linguistic variables when the journey theme is considered in isolation

	variable	freq F	freq R		% change	LL	Cohen's d	score
1.	us	2200	1419	+	54.97	457	0.42	1.55
2.	fpp	10615	6885	+	54.18	2160	0.53	1.54
3.	we	6363	4369	+	45.61	994	0.46	1.46
4.	your	9239	7317	+	26.27	584	0.61	1.26
5.	you	21475	17892	+	20.03	845	0.66	1.2
6.	JJS	3000	2585	+	16.06	80	0.43	1.16
7.	relcl	12789	11439	+	11.80	192	0.58	1.12
8.	<u>Negative</u>	32376	36280	-	10.76	549	0.43	0.89
9.	PRP\$	30733	27758	+	10.72	387	0.66	1.11
10.	WRB	8073	7307	+	10.48	97	0.52	1.10
11.	<u>i</u>	35177	38963	-	9.72	479	0.42	0.90
12.	Joy	26632	24280	+	9.68	277	0.66	1.10
13.	<u>Sadness</u>	20722	22820	-	9.20	251	0.43	0.91
14.	CCONJ	44685	41163	+	8.56	368	0.66	1.09
15.	unique words	586570	640533	-	8.42	5894	0.80	0.92
16.	conj	46682	43252	+	7.93	333	0.65	1.08
17.	Positive	48549	45136	+	7.56	316	0.68	1.08
18.	poss	30134	28089	+	7.28	183	0.61	1.07
19.	VCN	21443	23057	-	7.00	146	0.46	0.93
20.	prep	83904	78424	+	6.99	470	0.68	1.07
21.	IN	108063	101153	+	6.83	580	0.67	1.07
22.	pobj	76186	71315	+	6.83	409	0.67	1.07
23.	ADP	99155	92873	+	6.76	522	0.68	1.07
24.	TO	23034	21589	+	6.69	119	0.61	1.07
25.	punctuation	96115	90240	+	6.51	470	0.64	1.07
26.	<u>Fear</u>	24210	25843	-	6.32	133	0.46	0.94
27.	long_words	134573	126593	+	6.30	619	0.67	1.06
28.	pcomp	6904	6511	+	6.04	29	0.47	1.06
29.	VB	60899	57609	+	5.71	232	0.69	1.06
30.	DET	81916	77514	+	5.68	308	0.64	1.06
31.	DT	75471	71442	+	5.64	280	0.64	1.06
32.	mark	21019	19930	+	5.47	73	0.54	1.05
33.	VBZ	27669	29171	-	5.15	99	0.54	0.95
34.	Trust	32665	31074	+	5.12	101	0.65	1.05
35.	advcl	24425	23336	+	4.66	63	0.59	1.05
36.	attr	12404	11883	+	4.38	28	0.56	1.04
37.	SCONJ	16459	15793	+	4.22	35	0.5	1.04
38.	MD	19515	18761	+	4.02	38	0.59	1.04
39.	it	15393	16024	-	3.94	32	0.48	0.96

40.	Anticipation	34885	33593	+	3.85	62	0.67	1.04
41.	<u>Anger</u>	11761	12205	-	3.64	21	0.44	0.96
42.	ccomp	36901	38215	-	3.44	58	0.56	0.97
43.	VBG	23218	24015	-	3.32	34	0.53	0.97
44.	polysyllabic words	57012	55219	+	3.25	72	0.58	1.03
45.	xcomp	18795	18234	+	3.08	21	0.58	1.03
46.	VBP	45778	47200	-	3.01	54	0.64	0.97
47.	AUX	68813	70869	-	2.90	76	0.61	0.97
48.	NN	134290	138256	-	2.87	145	0.59	0.97
49.	ADV	80501	78262	+	2.86	80	0.65	1.03
50.	nsubj	113588	116890	-	2.82	118	0.61	0.97
51.	det	63262	61680	+	2.56	51	0.59	1.03
52.	PRON	149795	146134	+	2.51	114	0.66	1.03
53.	advmod	78897	77071	+	2.37	54	0.65	1.02
54.	ADJ	68731	67165	+	2.33	46	0.67	1.02
55.	JJ	62559	61282	+	2.08	33	0.66	1.02
56.	NOUN	166720	170127	-	2.00	86	0.60	0.98
57.	aux	66734	67864	-	1.67	24	0.61	0.98
58.	monosyllabic words	803972	792769	+	1.41	198	0.67	1.01

4.5.2 rollercoaster metaphor theme

In this analysis the data was split based on the presence or absence of the *rollercoaster* metaphor theme. Corpus F consists of all the posts that have been counted as containing at least one instance of the *rollercoaster* theme. Corpus R consists of all posts that do not contain at least one instance of the *rollercoaster* theme. It can be seen in the first results table below that, as with *journey*, there are a greater range of pronouns used in conjunction with the *rollercoaster* metaphor theme in comparison to posts that do not include the theme. However, in contrast to *journey* the ratio of positive to negative emotion words is slightly decreased where *rollercoaster* is used.

Table 4-23 density of rollercoaster theme in cancer.net

	posts	tokens	average post length (tokens)	standard deviation of pronouns	ratio positive to negative emotion words
Corpus F rollercoaster>0	1,507 (0.58%)	305,121	202	2.46	1.22
Corpus R rollercoaster=0	258,734	28,217,123	109	1.66	1.25

The following table lists all the variables that were found to be significantly different in density between the two corpora, ranked according to the proportion of change between corpus F and corpus R. There are 36 significantly different variables between posts in which *rollercoaster* is and is not used, compared to 61 for *journey*, suggesting that there is less of a difference in language style associated with the *rollercoaster* than with the *journey* theme. It can be seen that the first-person plural pronouns *fpp*, and *we*, are the most increased in corpus F in comparison to corpus R, while the first-person singular pronoun *i*, is the most decreased pronoun, a language style that has been associated with better mental health. In comparison to *journey*, where the *Trust*, *Joy* and *Positive* variables were found to be significantly increased in corpus F, *Trust*, *Joy* and *Positive* are significantly decreased in conjunction with *rollercoaster*, while *Anger* is significantly increased.

Table 4-24 most changed linguistic variables between posts that do and do not contain rollercoaster

	variable	freq F	freq R		% change	LL	Cohen's d	score
1.	fpp	10904	7130	+	52.92	517	0.53	1.53
2.	we	6860	4509	+	52.12	318	0.48	1.52
3.	attr	13391	11922	+	12.33	53	0.63	1.12
4.	Anger	13634	12249	+	11.31	46	0.55	1.11
5.	it	17672	15945	+	10.83	54	0.63	1.11
6.	RP	8341	7566	+	10.24	23	0.49	1.10
7.	CCONJ	45592	41379	+	10.18	125	0.67	1.10
8.	pcomp	7197	6534	+	10.15	20	0.50	1.10

9.	prt	8200	7458	+	9.95	22	0.49	1.10
10.	VBZ	31548	29023	+	8.70	64	0.67	1.09
11.	conj	47126	43475	+	8.40	90	0.65	1.08
12.	i	35592	38684	-	7.99	77	0.47	0.92
13.	unique words	585047	634742	-	7.83	1207	0.79	0.92
14.	DT	77061	71555	+	7.70	125	0.67	1.08
15.	DET	83616	77654	+	7.68	135	0.66	1.08
16.	polysyllabic words	58944	55168	+	6.84	76	0.63	1.07
17.	neg	13116	14077	-	6.82	20	0.43	0.93
18.	VBG	25541	23959	+	6.60	31	0.61	1.07
19.	MD	17613	18755	-	6.09	21	0.48	0.94
20.	<u>Trust</u>	29228	31090	-	5.99	34	0.51	0.94
21.	dobj	54221	57632	-	5.92	62	0.56	0.94
22.	ADP	98272	93189	+	5.45	82	0.66	1.05
23.	punctuation	95303	90429	+	5.39	78	0.63	1.05
24.	<u>Joy</u>	23122	24353	-	5.06	19	0.50	0.95
25.	long_words	133055	126711	+	5.01	94	0.66	1.05
26.	prep	82600	78713	+	4.94	57	0.65	1.05
27.	det	64742	61756	+	4.83	43	0.62	1.05
28.	ADV	82017	78296	+	4.75	53	0.68	1.05
29.	pobj	74531	71579	+	4.12	36	0.63	1.04
30.	<u>Positive</u>	43416	45247	-	4.05	23	0.54	0.96
31.	IN	105283	101474	+	3.75	43	0.64	1.04
32.	VBP	45267	46971	-	3.63	19	0.65	0.96
33.	advmod	79713	77063	+	3.44	27	0.67	1.03
34.	NN	133570	138133	-	3.30	46	0.58	0.97
35.	NOUN	165915	169947	-	2.37	29	0.59	0.98
36.	monosyllabic words	807627	792553	+	1.90	86	0.68	1.02

4.5.2.1 rollercoaster theme used in isolation

In the following analysis, to consider further the specific work of the *rollercoaster* theme, posts were excluded in which *rollercoaster* is used alongside other themes, such that corpus F consists of posts that only contain the *rollercoaster* theme, while corpus R consists of posts that contain none of the metaphor themes. In this case, there are 33 significantly different variables. However, unlike for *journey* there are no notable differences when *rollercoaster* is considered in isolation from the other themes.

Table 4-25 most changed linguistic variables when the rollercoaster theme is considered in isolation

	variable	freq F	freq R		% change	LL	Cohen's d	score
1.	fpp	10044	6885	+	45.88	291	0.49	1.46
2.	we	6343	4369	+	45.16	179	0.46	1.45
3.	attr	13487	11883	+	13.49	48	0.63	1.13
4.	Anger	13637	12205	+	11.73	37	0.52	1.12
5.	pcomp	7217	6511	+	10.85	17	0.48	1.11
6.	it	17580	16024	+	9.71	34	0.59	1.1
7.	CCONJ	45161	41163	+	9.71	87	0.65	1.1
8.	VBZ	31627	29171	+	8.42	46	0.65	1.08
9.	conj	46677	43252	+	7.92	61	0.63	1.08
10.	polysyllabic words	59574	55219	+	7.89	77	0.64	1.08
11.	<u>Joy</u>	22557	24280	-	7.10	29	0.47	0.93
12.	unique_words	597392	640533	-	6.74	684	0.77	0.93
13.	dobj	53934	57687	-	6.51	57	0.54	0.93
14.	MD	17541	18761	-	6.50	19	0.45	0.93
15.	DT	75956	71442	+	6.32	64	0.65	1.06
16.	<u>Trust</u>	29110	31074	-	6.32	29	0.50	0.94
17.	DET	82337	77514	+	6.22	68	0.65	1.06
18.	VBG	25478	24015	+	6.09	20	0.58	1.06
19.	long_words	133737	126593	+	5.64	91	0.66	1.06
20.	ADP	97638	92873	+	5.13	55	0.65	1.05
21.	prep	82303	78424	+	4.95	43	0.64	1.05
22.	punctuation	94665	90240	+	4.90	49	0.61	1.05
23.	ADV	82066	78262	+	4.86	42	0.65	1.05
24.	<u>Positive</u>	42993	45136	-	4.75	24	0.52	0.95
25.	<u>i</u>	37245	38963	-	4.41	18	0.48	0.96
26.	pobj	74366	71315	+	4.28	30	0.62	1.04
27.	det	64180	61680	+	4.05	23	0.60	1.04
28.	IN	104941	101153	+	3.75	32	0.63	1.04
29.	advmod	79726	77071	+	3.44	21	0.64	1.03
30.	NN	133845	138256	-	3.19	33	0.57	0.97
31.	RB	84268	81887	+	2.91	16	0.64	1.03
32.	NOUN	166092	170127	-	2.37	22	0.58	0.98
33.	monosyllabic words	808081	792769	+	1.93	68	0.66	1.02

4.5.3 fight metaphor theme

For the analysis in this section the data was split based on the presence or absence of the *fight* metaphor theme. Corpus F consists of all the posts that have been counted as containing at least one instance of the *fight* metaphor theme, while corpus R consists of all the posts that do not contain at least one instance of the *fight* theme. It can be seen from the first results table below that a greater range of pronouns was found to be used in conjunction with the *fight* theme compared to posts that do not contain the theme. However, unlike with the *journey* theme, but the same as for the *rollercoaster* theme, the ratio of positive to negative emotion words was found to be slightly decreased where the *fight* theme is used.

Table 4-26 density of fight theme in cancer.net

	posts	tokens	average post length (tokens)	standard deviation of pronouns	ratio positive to negative emotion words
Corpus F fight>0	1,945 (0.75%)	446,149	229	3.07	1.21
Corpus R fight=0	258,296	28,073,613	109	1.65	1.25

The following table lists variables with a significantly different density between corpus F and Corpus R; there are 46 such variables associated with the *fight* metaphor, compared with 61 for the *journey* condition, and 36 for the *rollercoaster* condition. It can be seen that the first-person plural pronouns *fpp*, and *we*, are the most increased in corpus F in comparison to corpus R, while *you*, and *your* are the most decreased key variables. The negative emotion variables *Anger*, *Disgust*, *Fear*, and *Sadness* are increased where *fight* is used, while the positive emotion variable *Anticipation* is decreased.

Table 4-27 most changed linguistic variables between posts that do and do not contain the fight metaphor theme

	variable	freq F	freq R		% change	LL	Cohen's d	score
1.	fpp	10375	7120	+	45.71	571	0.52	1.46
2.	we	6547	4503	+	45.40	356	0.48	1.45
3.	VBD	43550	36018	+	20.91	647	0.48	1.21
4.	Anger	14688	12227	+	20.13	204	0.67	1.20
5.	Disgust	11097	9327	+	18.98	139	0.61	1.19
6.	PDT	2197	1906	+	15.23	18	0.42	1.15
7.	unique words	553537	635548	-	12.90	4865	0.84	0.87
8.	RP	8497	7560	+	12.39	49	0.59	1.12
9.	prt	8363	7452	+	12.22	47	0.58	1.12
10.	PRP\$	31227	27944	+	11.75	163	0.75	1.12
11.	me	6202	5553	+	11.68	32	0.46	1.12
12.	<u>you</u>	16026	18015	-	11.04	100	0.47	0.89
13.	WP	5509	4992	+	10.36	23	0.53	1.10
14.	poss	31196	28293	+	10.26	126	0.72	1.10
15.	pcomp	7193	6532	+	10.12	28	0.55	1.10
16.	VBP	42410	47029	-	9.82	206	0.67	0.90
17.	Fear	28298	25782	+	9.76	104	0.65	1.10
18.	Sadness	24936	22736	+	9.67	91	0.64	1.10
19.	relcl	12579	11531	+	9.08	41	0.61	1.09
20.	CCONJ	45115	41370	+	9.05	145	0.73	1.09
21.	<u>your</u>	6776	7386	-	8.26	23	0.44	0.92
22.	PRON	155910	146231	+	6.62	275	0.76	1.07
23.	PRP	112631	106269	+	5.99	164	0.73	1.06
24.	NNS	34789	36999	-	5.97	59	0.58	0.94
25.	my	14267	13553	+	5.26	16	0.54	1.05
26.	ADP	97927	93175	+	5.10	105	0.73	1.05
27.	neg	14760	14057	+	5.00	15	0.61	1.05
28.	<u>Anticipation</u>	32063	33586	-	4.53	31	0.65	0.95
29.	prep	82248	78705	+	4.50	69	0.71	1.05
30.	amod	36208	37891	-	4.44	33	0.60	0.96
31.	<u>i</u>	36972	38681	-	4.42	34	0.54	0.96
32.	Negative	37561	36070	+	4.13	27	0.63	1.04
33.	TO	22522	21644	+	4.06	15	0.66	1.04
34.	pobj	74417	71570	+	3.98	49	0.70	1.04
35.	punctuation	94025	90436	+	3.97	62	0.63	1.04
36.	polysyllabic words	53081	55247	-	3.92	38	0.57	0.96
37.	NOUN	163656	170002	-	3.73	105	0.65	0.96

38.	IN	105097	101464	+	3.58	56	0.71	1.04
39.	NN	133527	138156	-	3.35	69	0.65	0.97
40.	JJ	59312	61267	-	3.19	28	0.67	0.97
41.	DET	80011	77690	+	2.99	30	0.67	1.03
42.	ADJ	65229	67154	-	2.87	24	0.68	0.97
43.	monosyllabic words	812309	792473	+	2.50	216	0.75	1.03
44.	long_words	123748	126838	-	2.44	33	0.63	0.98
45.	DT	73316	71593	+	2.41	18	0.67	1.02
46.	VERB	167948	164999	+	1.79	23	0.74	1.02

4.5.3.1 fight theme used in isolation

In the following analysis, to consider further the specific work of the *fight* theme, posts were excluded in which *fight* is used alongside other themes, such that corpus F contains posts that only contain the *fight* theme, while corpus R contains posts that contain none of the metaphor themes. The most notable difference when *fight* was considered separately from the other themes is that the prevalence of the first-person singular pronoun *i* was not found to be significantly different between corpus F and corpus R, suggesting that there is no particular relationship between *i* and the *fight* metaphor used alone: *i* may be used with *fight* in terms of an ideological dilemma, potentially related to negotiating subject positions in conjunction with the experience and characterisation of cancer, and participation on the forum.

Table 4-28 most changed linguistic variables when the fight theme is considered in isolation

	variable	freq F	freq R		% change	LL	Cohen's d	score
1.	we	6265	4369	+	43.39	252	0.49	1.43
2.	fpp	9858	6885	+	43.18	394	0.53	1.43
3.	Anger	14706	12205	+	20.49	167	0.66	1.2
4.	VBD	43148	36019	+	19.79	462	0.47	1.2
5.	Disgust	11162	9340	+	19.51	116	0.60	1.2
6.	PDT	2252	1886	+	19.40	23	0.42	1.19
7.	me	6308	5563	+	13.40	33	0.45	1.13
8.	PRP\$	31406	27758	+	13.15	160	0.74	1.13
9.	unique words	563954	640533	-	11.96	3331	0.82	0.88

10.	RP	8418	7570	+	11.20	32	0.56	1.11
11.	Fear	28733	25843	+	11.18	109	0.65	1.11
12.	poss	31228	28089	+	11.18	118	0.71	1.11
13.	prt	8260	7462	+	10.68	29	0.55	1.11
14.	relcl	12568	11439	+	9.86	38	0.58	1.1
15.	Sadness	25045	22820	+	9.75	73	0.63	1.1
16.	<u>you</u>	16256	17892	-	9.14	54	0.45	0.91
17.	my	14737	13585	+	8.48	33	0.54	1.08
18.	PRON	157965	146134	+	8.10	325	0.75	1.08
19.	CCONJ	44475	41163	+	8.04	90	0.71	1.08
20.	VBP	43437	47200	-	7.97	108	0.64	0.92
21.	<u>your</u>	6735	7317	-	7.96	17	0.41	0.92
22.	PRP	114537	106345	+	7.70	215	0.72	1.08
23.	conj	46574	43252	+	7.68	87	0.70	1.08
24.	NNS	34342	37068	-	7.35	72	0.56	0.93
25.	amod	35412	37836	-	6.41	55	0.56	0.94
26.	neg	14963	14154	+	5.72	16	0.58	1.06
27.	polysyllabic words	52398	55219	-	5.11	51	0.54	0.95
28.	TO	22680	21589	+	5.05	19	0.65	1.05
29.	NOUN	161849	170127	-	4.87	143	0.63	0.95
30.	Anticipation	32009	33593	-	4.72	26	0.63	0.95
31.	ccomp	39918	38215	+	4.46	26	0.66	1.04
32.	NN	132247	138256	-	4.35	92	0.64	0.96
33.	ADP	96788	92873	+	4.22	57	0.71	1.04
34.	JJ	58790	61282	-	4.07	36	0.65	0.96
35.	Negative	37717	36280	+	3.96	20	0.62	1.04
36.	long_words	121781	126593	-	3.80	65	0.60	0.96
37.	ADJ	64781	67165	-	3.55	30	0.67	0.96
38.	prep	81106	78424	+	3.42	32	0.70	1.03
39.	det	59769	61680	-	3.10	21	0.59	0.97
40.	pobj	73434	71315	+	2.97	22	0.70	1.03
41.	IN	104015	101153	+	2.83	28	0.69	1.03
42.	monosyllabic words	815021	792769	+	2.81	216	0.74	1.03
43.	ADV	80172	78262	+	2.44	16	0.68	1.02
44.	VERB	169073	165196	+	2.35	31	0.72	1.02
45.	nsubj	119581	116890	+	2.30	21	0.69	1.02
46.	punctuation	92228	90240	+	2.20	15	0.62	1.02

4.5.4 battle

For the analysis in this section the data was split based on the presence or absence of the *battle* metaphor theme. Corpus F consists of all the posts that have been counted as containing at least one instance of the *battle* metaphor theme, while corpus R consists of all posts that do not contain at least one instance of the *battle* theme. It can be seen from the first results table below that as with all the metaphor themes there is a greater range of pronouns used where *battle* is present compared to posts where it is not, with *battle* having the highest standard deviation of pronouns of all the metaphoric conditions. However, as with *rollercoaster* and *fight* the ratio of positive to negative emotion words is slightly decreased in posts in which *battle* is present.

Table 4-29 density of battle theme in cancer.net

	posts	tokens	average post length (tokens)	standard deviation of pronouns	ratio positive to negative emotion words
Corpus F fight>0	2,004 (0.77%)	446,227	223	3.89	1.20
Corpus R fight=0	258,237	28,073,524	109	1.65	1.25

The following table lists variables with a significantly different density between corpus F and Corpus R; there are 36 such variables associated with the *battle* metaphor, compared to 61 for the *journey* condition, 36 for the *rollercoaster* condition, and 46 for the *fight* condition. It can be seen that the first-person plural pronouns *we*, and *fpp*, are the most increased variables in corpus F in comparison to corpus R, and *you*, and *i* are the most decreased pronouns. In terms of emotion variables, *Anger*, *Disgust*, *Sadness*, and *Fear*, are significantly increased, and the composite emotion variables *Negative* and *Positive* are also significantly increased where *battle* is used; there are no significantly decreased emotion variables.

Table 4-30 most changed linguistic variables between posts that do and do not contain the battle metaphor theme

	variable	freq F	freq R		% change	LL	Cohen's d	score
1.	we	6096	4510	+	35.15	220	0.43	1.35
2.	fpp	9354	7137	+	31.07	275	0.45	1.31
3.	Anger	15481	12214	+	26.75	353	0.71	1.27
4.	Disgust	11676	9317	+	25.31	242	0.66	1.25
5.	PDT	2268	1905	+	19.04	29	0.43	1.19
6.	VBD	41880	36045	+	16.19	394	0.47	1.16
7.	PRP\$	32418	27922	+	16.10	302	0.75	1.16
8.	poss	32717	28265	+	15.75	293	0.73	1.16
9.	Sadness	25926	22721	+	14.11	190	0.67	1.14
10.	Fear	29048	25769	+	12.72	176	0.67	1.13
11.	my	15145	13539	+	11.86	80	0.56	1.12
12.	me	6194	5554	+	11.53	31	0.43	1.12
13.	unique words	563056	635398	-	11.39	3766	0.84	0.89
14.	<u>you</u>	15967	18016	-	11.37	106	0.41	0.89
15.	relcl	12832	11529	+	11.30	62	0.61	1.11
16.	conj	48296	43442	+	11.17	230	0.71	1.11
17.	CCONJ	45132	41369	+	9.10	146	0.71	1.09
18.	VBP	43131	47016	-	8.26	145	0.66	0.92
19.	it	14667	15986	-	8.25	49	0.49	0.92
20.	pcomp	7023	6534	+	7.48	16	0.52	1.07
21.	Negative	38673	36053	+	7.27	82	0.64	1.07
22.	pobj	76703	71533	+	7.23	160	0.69	1.07
23.	prep	84202	78672	+	7.03	167	0.70	1.07
24.	VBZ	27123	29083	-	6.74	59	0.58	0.93
25.	ADP	98898	93160	+	6.16	152	0.69	1.06
26.	IN	107324	101429	+	5.81	148	0.70	1.06
27.	TO	22802	21640	+	5.37	27	0.64	1.05
28.	PRON	153610	146268	+	5.02	159	0.74	1.05
29.	NNS	35363	36988	-	4.39	32	0.57	0.96
30.	<u>i</u>	37140	38679	-	3.98	27	0.56	0.96
31.	Positive	46512	45211	+	2.88	16	0.64	1.03
32.	AUX	68660	70654	-	2.82	25	0.66	0.97
33.	PRP	109144	106327	+	2.65	32	0.70	1.03
34.	DET	79419	77698	+	2.21	17	0.63	1.02
35.	NOUN	167153	169944	-	1.64	20	0.63	0.98
36.	monosyllabic words	802919	792622	+	1.30	58	0.71	1.01

4.5.4.1 Battle theme used in isolation

In the following analysis, to consider further the specific work of the *battle* theme, posts were excluded in which *battle* is used alongside other themes, such that corpus F consists of posts that only contain the *battle* theme, while corpus R consists of posts that contain none of the metaphor themes. It can be seen that, as with the *fight* metaphor considered separately, when *battle* is considered in isolation from the other metaphor themes the most notable difference is that the first-person singular pronoun *i* is not significantly reduced in density between corpus F and corpus R.

Table 4-31 most changed linguistic variables when the battle theme is considered in isolation

	variable	freq F	freq R		% change	LL	Cohen's d	score
1.	we	5838	4369	+	33.60	154	0.45	1.34
2.	fpp	8888	6885	+	29.09	184	0.46	1.29
3.	Anger	15606	12205	+	27.86	300	0.72	1.28
4.	Disgust	11892	9340	+	27.33	221	0.67	1.27
5.	PDT	2298	1886	+	21.88	29	0.42	1.22
6.	PRP\$	32650	27758	+	17.63	281	0.73	1.18
7.	my	15898	13585	+	17.02	129	0.56	1.17
8.	poss	32739	28089	+	16.55	252	0.71	1.17
9.	VBD	41892	36019	+	16.31	314	0.46	1.16
10.	Sadness	26198	22820	+	14.80	164	0.69	1.15
11.	Fear	29588	25843	+	14.49	179	0.68	1.14
12.	me	6292	5563	+	13.12	32	0.41	1.13
13.	relcl	12719	11439	+	11.18	48	0.59	1.11
14.	conj	48021	43252	+	11.03	175	0.71	1.11
15.	unique_words	575174	640533	-	10.20	2386	0.83	0.9
16.	CCONJ	44974	41163	+	9.26	118	0.71	1.09
17.	it	14603	16024	-	8.87	45	0.46	0.91
18.	Negative	39439	36280	+	8.71	92	0.65	1.09
19.	VBP	43942	47200	-	6.90	79	0.64	0.93
20.	pobj	76052	71315	+	6.64	106	0.70	1.07
21.	PRON	155674	146134	+	6.53	210	0.73	1.07
22.	prep	83467	78424	+	6.43	110	0.71	1.06
23.	VBZ	27304	29171	-	6.40	42	0.56	0.94
24.	NNS	34834	37068	-	6.03	47	0.55	0.94
25.	TO	22870	21589	+	5.93	26	0.64	1.06

26.	ADP	98053	92873	+	5.58	98	0.70	1.06
27.	IN	106377	101153	+	5.16	92	0.70	1.05
28.	PRP	110891	106345	+	4.27	66	0.69	1.04
29.	ADV	81283	78262	+	3.86	40	0.70	1.04
30.	NOUN	165371	170127	-	2.8	46	0.64	0.97
31.	RB	83939	81887	+	2.51	18	0.69	1.03
32.	AUX	69093	70869	-	2.51	15	0.66	0.97
33.	NN	135548	138256	-	1.96	18	0.65	0.98
34.	monosyllabic words	805230	792769	+	1.57	67	0.72	1.02

4.5.5 Section summary: comparing language style between metaphoric conditions

For each metaphoric condition, first-person plural pronouns were the most increased variables in Corpus F compared to Corpus R, with *fpp*, and *we* significantly increased in each condition, while at the same time, the first-person singular pronoun *i* was significantly decreased in each condition. This is a language style that has been associated with better mental health, and in each case the effect was strongest for the *journey* theme. In addition, a wider range of pronouns was used in each metaphoric condition in comparison to the associated non-metaphoric condition, which is another language style that has been associated with better mental health. However, the ratio of positive to negative emotion words was found to be increased in the *journey* metaphoric condition only; for all other metaphoric conditions the ratio was decreased: more negative than positive emotion words are used where the *rollercoaster*, *fight*, and *battle* metaphor themes are used in comparison to the posts in which they are not used.

The following table summarises the associations that have been found between the emotion variables and the metaphor themes. These results suggest that the interpretative repertoire of which the *journey* theme is a part supports discussion of positive emotions and suppresses discussion of negative emotions. As was suggested in chapter three, the *rollercoaster* theme does not appear particularly to support wider discussion, for example

around emotions: it appears to be a relatively one-dimensional metaphor based on the concept of sudden ups and downs, and this was also supported by the finding in the supervised learning investigation earlier in this chapter that *rollercoaster* is not characterised by particular use of pronouns. The *fight* and *battle* themes both appear to support discussion predominantly of negative, but not positive, emotion. It is notable that none of the metaphor themes is associated with *Surprise*.

Table 4-32 association of emotion variables with metaphor themes

	Positive	Anticipation	Joy	Surprise	Trust	Anger	Disgust	Fear	Sadness	Negative
journey	+	+	+		+	-		-	-	-
rollercoaster	-		-		-	+				
fight		-				+	+	+	+	+
battle	+					+	+	+	+	+

An analysis was also made in which only the specific metaphor theme under investigation was included, such that all posts that contain any other metaphor theme, even if they also include the theme under investigation, were excluded. In this case, each of the metaphor themes was found to have the key variables *we*, and *fpp*, as the most increased in comparison to the reference corpus, except for *journey* which had the three key variables *us*, *we*, and *fpp*, as the most increased, suggesting that the pronoun *us* is of high significance in conjunction with *journey*. It is of particular note that *our* was found to be the most increased variable for *journey*, however when *journey* was considered in isolation from the other themes density of *our* was not significantly different to its density in the reference corpus. This strongly suggests that *our* in conjunction with *journey* is used to support an ideological dilemma relating to subject positions, and to community rhetoric.

Use of the key variable *i* was found to be decreased only for the *journey* and *rollercoaster* themes used in isolation, and not for the *fight* and *battle* themes. In terms of subject positions, these findings suggest that participants using the *fight* and *battle* themes in

isolation are still representing and forming their own identity as a legitimate cancer.net forum participant, which entails use of the first-person singular pronoun *i*. This is also supported by the fact that the range of personal pronouns used was highest for *battle*, and second highest for *fight*. It is also consistent with the results from chapter three in which *battle* was found to be more used by newer participants than by more established participants, while *fight* was found to be equally used by both groups. However, for each theme considered in isolation, the increased use of the first-person plural pronoun *we* shows that where any of the metaphor themes is used there is a sense of membership entitlement, and a focus away from the self alone.

In summary, the most significant change in language towards a style associated with better mental health was found in the condition in which *journey* metaphors are present in a post. There are also a greater number of significantly different linguistic variables for *journey* in comparison to the other themes. These results are consistent with the supervised learning analysis above, in which *journey* was predicted most accurately based on surrounding language style.

4.6 Chapter discussion

In this chapter supervised learning techniques were used to explore the organisation of the metaphor themes based on their surrounding language style, and to further consider the approach set out in chapter two, that since metaphoric nouns may be the most active expressions of metaphor, they are likely to be surrounded by a more distinctive language style. The validity of the organisation of the metaphor themes was confirmed, as was the focus on metaphoric nouns, which for the TEXT method were found have a stronger effect on surrounding language style than metaphoric nouns and verbs considered together, in that that language style was more detectable by the supervised learning

models. However, for the COUNT method *battle* was more accurately predicted when metaphor instantiated as both nouns and verbs was included.

In the more finely grained comparison of corpora that do and do not contain each metaphor theme, many significant differences were found between corpus F (posts that contain the metaphor theme under investigation), and corpus R (posts that do not contain the metaphor theme under investigation). Although all of the metaphor themes were associated with increased use of first-person plural pronouns and decreased use of the first-person singular pronoun *i*, this was most distinct for the *journey* theme, and least distinct for the *battle* theme. In addition, posts containing the *journey* theme were found to have a higher ratio of positive to negative emotion words than posts not containing the theme, while this was not the case for the other metaphor themes. However, in the analysis in which each metaphor theme was considered in isolation from the other themes, although the first-person plural pronouns *we* and *fpp* were increased for every metaphoric condition, for the *fight* and *battle* themes used alone there was no decrease in use of *i*; only the *journey* and *rollercoaster* themes used in isolation were associated with decreased use of *i*. In addition, while there was an increase in the standard deviation of pronouns for all the metaphoric conditions, this was highest for the *battle* condition. This suggests, as discussed above, that the *battle* theme, which is used more by newer participants than by more established participants, is associated with a negotiation and representation of the participant's own subject position in the context of the established wider community. It is also notable that *our* was found to be significantly increased in conjunction with the *journey* theme, but not when *journey* is used in isolation, strongly suggesting that *our* in conjunction with *journey* plays a particular role in an ideological dilemma to negotiate subject positions and promote community rhetoric.

Although in the current chapter significant differences in language style were found to be associated with use of particular themes, the data was considered as a single time period when it spans more than 12 years. In the next chapter, consideration is given to how use of the different metaphor themes changes over time, and how the different language variables representing particular language styles are associated with them over time.

5 DIACHRONIC ANALYSIS

5.1 Introduction

In chapter four it was shown that each of the metaphor themes in use on cancer.net, *journey*, *rollercoaster*, *fight*, and *battle*, is surrounded by a distinctive language style such that its presence in any post can be predicted with high accuracy by a supervised learning model based solely on the language surrounding it. In the current chapter consideration is given to how use of the different metaphor themes changes over time, and how the different language variables representing particular language styles are associated with them over time.

In the diachronic analysis of the current chapter each of the 153 months of cancer.net data is considered as a separate data point. The Python SciPy *linregress* method is used to calculate least squares regression, and the interpretation of *r* is based on the standard described in the table below.

**Table 5-1 interpretation of *r*
absolute value**

absolute value	relationship between variables
0	none
≥ 0.3	weak
≥ 0.5	moderate
≥ 0.7	strong
1	exact linear

Tay (2019) notes that the issue of autocorrelation, which relates to the presence of internal relationships within temporally ordered observations, is usually overlooked in corpus linguistic research, in which independence between time periods is assumed, rather than possible interdependence between temporally ordered observations. Tay suggests time series analysis as a complementary methodological approach that can address such limitations. Time series analysis may "offer a new type of structural insight

into discourse behavior over time", and "the possibility of forecasting future values" (Tay, 2019, p. 11). Time series analysis is not undertaken in the current study, which focuses on trends across the whole data. However, plots of the diachronically focused analysis carried out here help to address this issue, allowing the visual inspection which has a fundamental role in time series analysis (Koplenig, 2017).

5.1.1 Organisation of the chapter

In the next section the density of the metaphor themes over time in cancer.net is considered, to understand how their use evolves on the forum. Following that, an analysis is made of the correlation of each metaphor theme with key variables associated with mental health over time to consider if and how particular language styles come to become associated with particular themes. Finally, consideration is given to the change in density of metaphor themes and key variables over time for comparative groups of participants who are more established, or less established, in the cancer.net community. At the end of this chapter cancer.net results from the current and previous chapters are discussed as a whole.

5.2 Density of metaphor use over time

In this section the density of use of each metaphor theme over time is explored, with themes that are adopted and adapted to characterise this community and perform particular work within it expected to increase in prevalence since the start of the cancer.net forum.

5.2.1 Mean density of each metaphor theme per post, by month

The mean density of a metaphor theme per post in any month is the mean of instances of that metaphor in a post, divided by the number of tokens in that post. In the following

heatmaps, the first of which is unsmoothed and the second smoothed with sigma=2, the mean density of each metaphor theme per post is plotted for each month. The scale at the right of the plot shows that higher density in any month is represented as brighter yellow, while zero use of a metaphor theme is represented as dark blue. It can be seen from the plot that *journey* becomes more densely used per post over time, with minimal use towards the start of the forum, while *fight* and *battle* appear to become less used over time.

Figure 5-1 mean density of each metaphor theme per post, by month

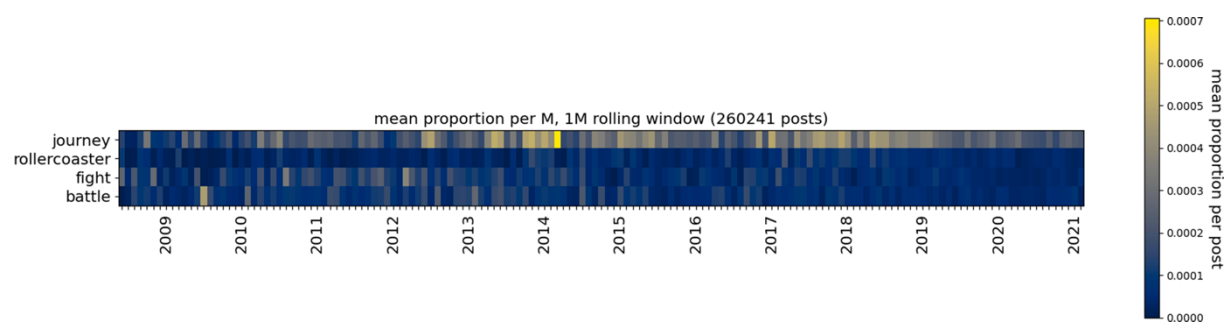
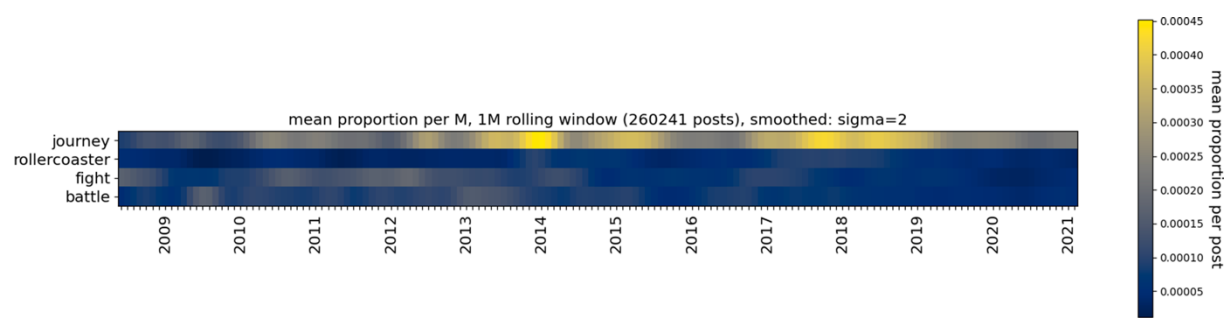


Figure 5-2 mean density of each metaphor theme per post, by month, smoothed, sigma=2



The following table shows the minimum, maximum, mean, and standard deviation of the mean density of use of each metaphor theme per post per month. For all themes the minimum value is 0, showing that each of the themes is absent from cancer.net in at least one month. The mean values show that *journey* is the most prevalent theme in the data. In addition, density of use of *journey* per post shows a weak increase over time ($r=0.3356$), while density of use of *fight* per post shows a weak decrease over time ($r=-0.4274$).

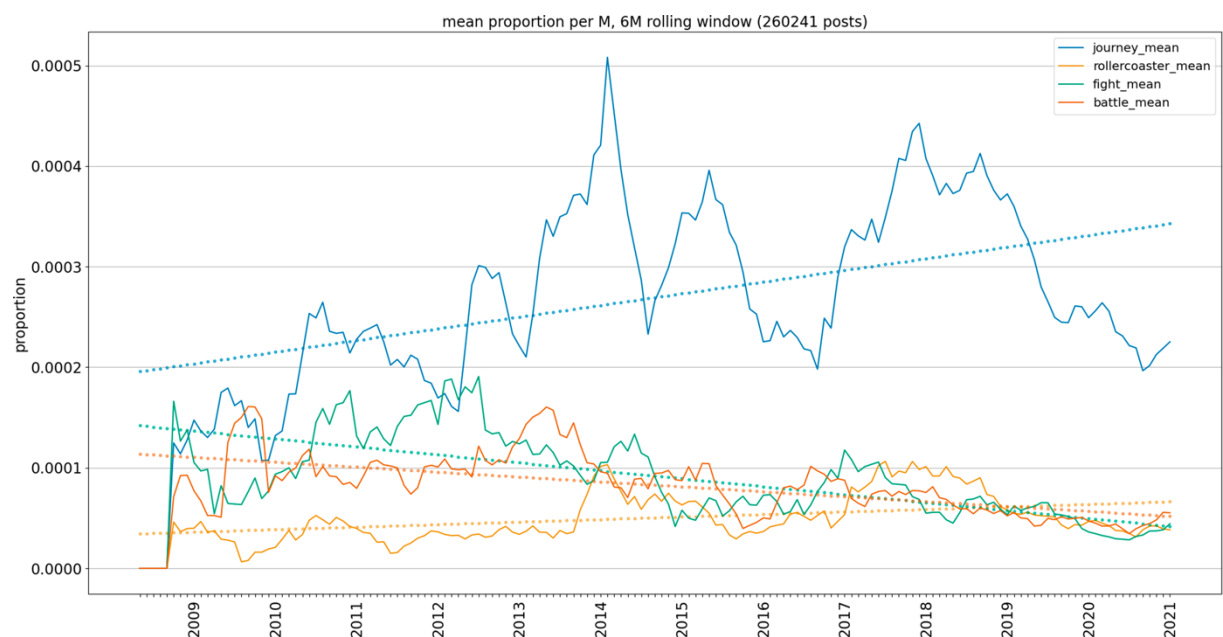
Density of use of *rollercoaster* per post also increases over time ($r=0.2262$) and density of use of *battle* per post decreases ($r=-0.2887$), but although these changes are notable, they do not meet the 0.3 interpretation threshold.

Table 5-2 mean density of each metaphor theme per post, over time

theme	min	max	mean	std	r	slope
journey	0.0	0.0007	0.0003	0.0001	0.3356	1e-06
rollercoaster	0.0	0.0002	0.0001	0.0000	0.2262	2e-07
fight	0.0	0.0004	0.0001	0.0001	-0.4274	-7e-07
battle	0.0	0.0005	0.0001	0.0001	-0.2887	-4e-07

To show the pattern of change over time in a different format, the same data was plotted as a line graph, below. The *journey* line is blue, *rollercoaster* is yellow, *fight* is green, and *battle* is orange. A rolling window of 6 months was used. A least squares regression line (dotted) was plotted for each metaphor theme with the slope described in the table above.

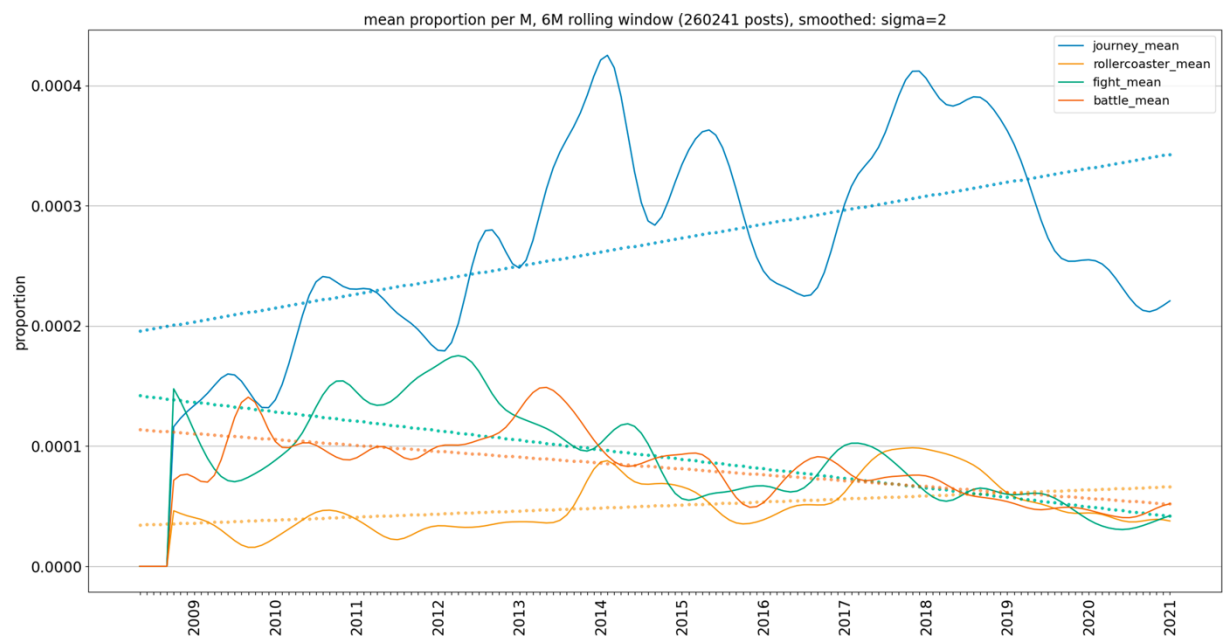
Figure 5-3 mean density of each metaphor theme per post, by month



In the following plot, the same data was smoothed using the *gaussian_filter1d* filter from the *scipy.ndimage.filters* Python library. Smoothing is a process in which data points are averaged with their neighbours in a series, to blur the sharp edges in the data. Smoothing

suppresses high frequency signal and enhances low frequency signal, such that data points are brought towards the middle. This can show the trend more clearly, but at the same time the extreme points of the data become less visible. For the following plot the smoothing factor sigma was set to two.

Figure 5-4 mean density of each metaphor theme per post, by month, smoothed, sigma=2



The smoothed plot shows more clearly the pattern of change of density of metaphor use per post, by month. It can be seen that most of the change happens in the first four years of the forum. After four years the pattern of metaphor use in this community is more established, such that *journey* has become and remains the dominant metaphor theme.

5.2.2 Proportion of posts using each metaphor theme by month

In comparison to the analysis in the previous section, which considered the mean density of each theme per post for each month, in this section the focus is on the proportion of posts containing a particular theme in any month. So, in this section simply the presence or absence of a metaphor theme in a post is considered, rather than its density of use per post.

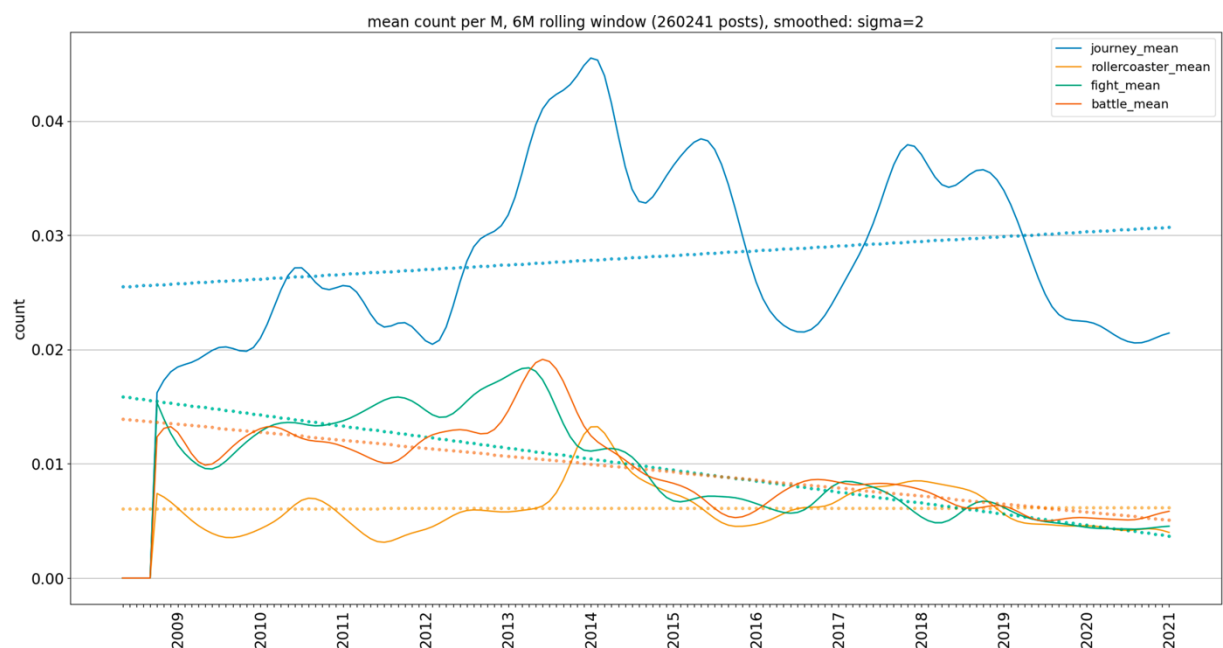
The following table shows that there is a moderate decrease in the proportion of posts using the *fight* ($r=-0.5859$) and *battle* ($r=-0.4905$) themes over time, while the slight increase in the number of posts using the *journey* theme per month is not significant.

Figure 5-5 proportion of posts containing each metaphor theme per month

	min	max	mean	std	r	slope
journey	0.0032	0.0652	0.0281	0.0108	0.1406	3.44e-05
rollercoaster	0.0	0.0221	0.0061	0.0040	0.0067	6e-07
fight	0.0	0.0395	0.0098	0.0061	-0.5859	-8.02e-05
battle	0.0	0.0294	0.0095	0.0053	-0.4905	-5.82e-05

The following plot shows the pattern of change of presence of each metaphor theme per post, by month, which is similar to the pattern found for the analysis in the previous section in which the density of use of each theme per post was considered.

Figure 5-6 proportion of posts using each metaphor theme per month



Another important consideration in analysing this pattern of change is the distribution of participants posting over time. Because if a greater number of new participants are posting in a particular period, it is to be expected that the proportion of posts using a community metaphor theme would decrease in that period, whereas if a higher proportion

of more established participants are posting, the proportion of posts using a community metaphor theme should increase.

The following plots, which use a log y axis, demonstrate that for each metaphor theme, the proportion of posts using that theme tends to fall away with an influx of new participants, and correspondingly, use of that theme per post tends to increase as the influx of new participants levels off (the number of posts in a month is the pink line; the number of unique participants posting per month is the brown line).

Figure 5-7 number of posts using each metaphor theme by month, 6M rolling window, smoothed, logy

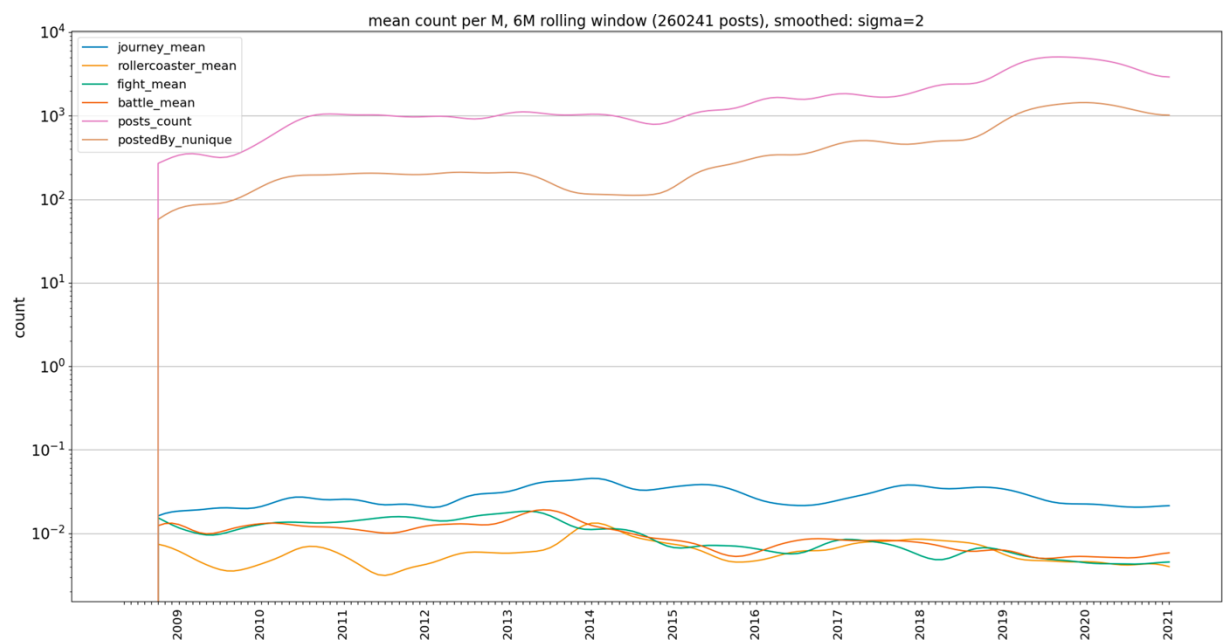
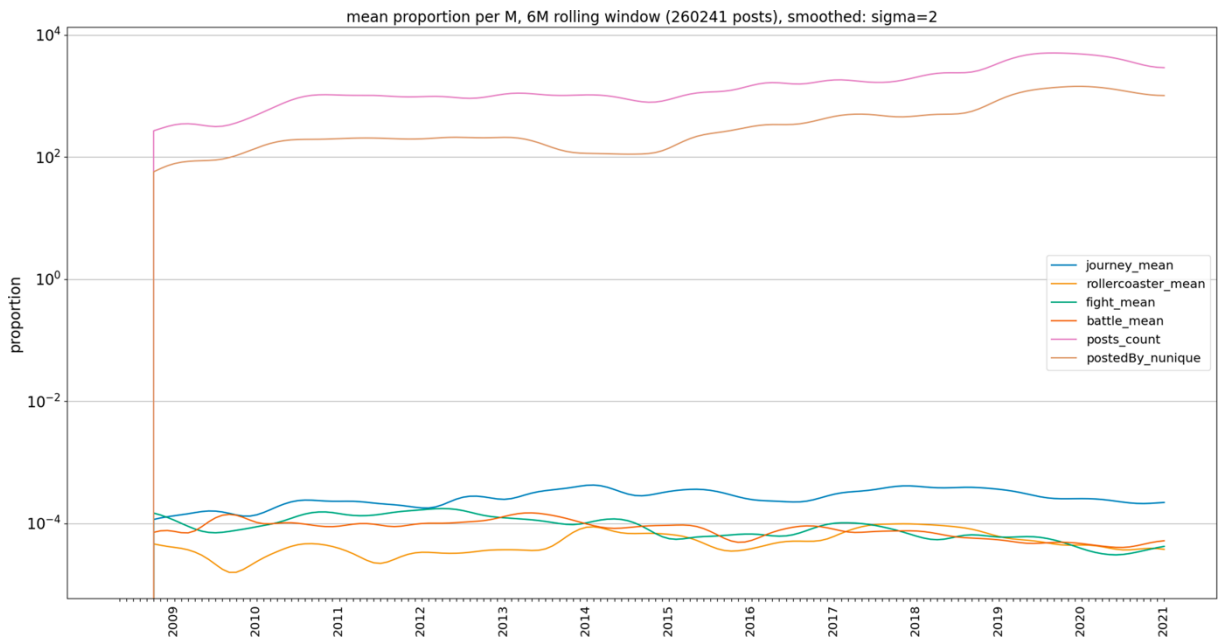


Figure 5-8 proportion of posts using each metaphor theme by month, 6M rolling window, smoothed, logy



While the least-squares regression analysis above suggested that *journey* has a particular role in this community in that its use in the community increases over time, the analysis in the current section suggests that regardless of whether they are community metaphors, use of all of the metaphor themes may be more associated with more established forum participants.

There may also be a longer-term effect in which if there is a large influx of new participants to the forum, the community metaphors may become relatively dis-established as their proportion relative to other posts decreases: they may become less powerful or meaningful as metaphors adapted for particular work in this community for a period. This may also open up an opportunity for renegotiation or adjustment of the work of the metaphor themes.

5.2.3 Section summary: density of metaphor use over time

In this section it has been shown that the *journey* metaphor theme becomes more dominant over time on cancer.net. Use of the *rollercoaster* theme does not change significantly over time, while use of the *fight* and *battle* themes decreases.

Table 5-3 change over time for metaphor themes based on density per post (left) and count (right)

	r (density)	slope (density)	r (count)	slope (count)
journey	0.3356	1e-06	0.1406	3.44e-05
rollercoaster	0.2262	2e-07	0.0067	6e-07
fight	-0.4274	-7e-07	-0.5859	-8.02e-05
battle	-0.2887	-4e-07	-0.4905	-5.82e-05

It can be seen from the summary table above that the density of use of *journey* per post is more increased over time ($r=0.3356$) than is the number of posts containing *journey* ($r=0.1406$), a pattern which is echoed by *rollercoaster*. The opposite pattern is true for *fight* and *battle*: their density of use per post decreases less over time than does the number of posts containing them. This pattern supports the view of *journey* and *rollercoaster* as community metaphor themes, in that they are increasingly used multiple times in a single post to address the community interpretative repertoires of which they are a part.

It was also shown that where a greater number of unique participants make posts in any month, suggesting an influx of newer participants, although the number of posts made in that month increases, as would be expected with more participants, at the same time use of the metaphor themes decreases. This supports the previous insight that the focus for newer participants is more literal.

5.3 Correlation of metaphor themes with key variables over time

The analysis in the current section again groups posts by month for comparison. For each metaphor theme consideration is given to the change in association between that theme and (i) the two most increased and two most decreased key variables (personal pronouns and emotion variables) associated with its use from chapter four and (ii) the eight emotion word variables *Anticipation*, *Joy*, *Surprise*, *Trust*, *Anger*, *Disgust*, *Fear*, and *Sadness*, and the meta variables *Positive* and *Negative* into which they are grouped, over time. This is done by calculating for each separate month the correlation of each of these variables with the metaphor theme under consideration.

5.3.1 journey metaphor theme

5.3.1.1 journey correlation over time with most changed key variables

It was shown in chapter four that *our* and *us* are the most increased key variables, and *i* and *Negative* the most decreased, when corpus F (posts in which *journey* is present) is compared with corpus R (posts in which *journey* is not present).

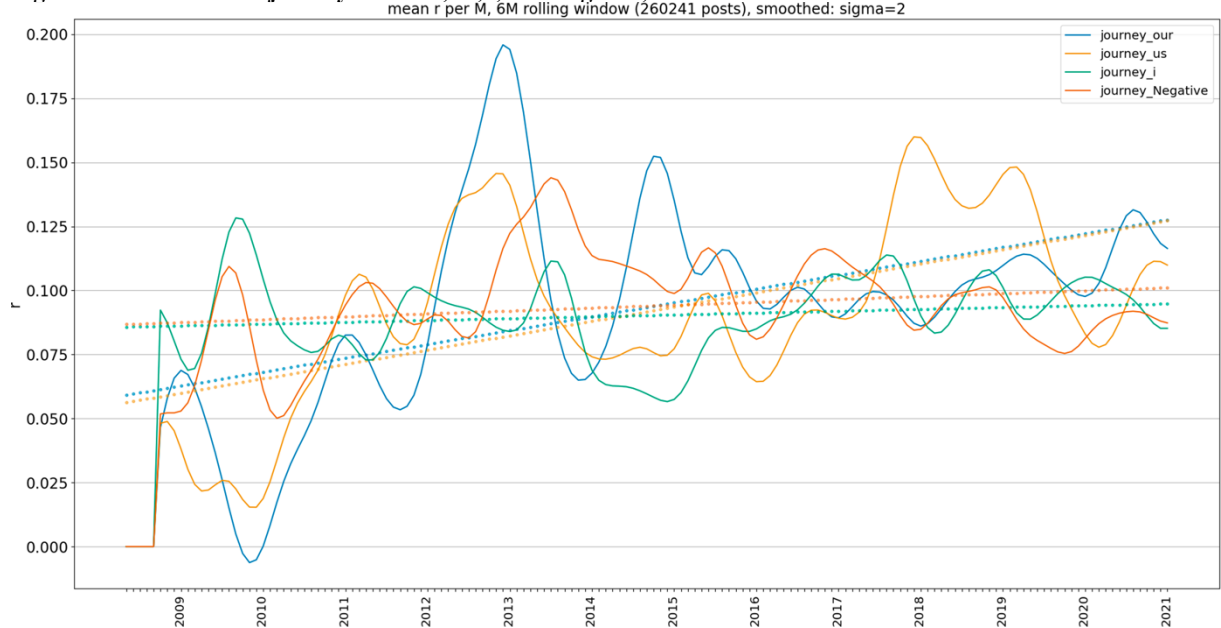
The following table shows the minimum, maximum, mean, and standard deviation of r for the correlation of *journey* with each of these most-changed key variables, and their change over time (slope). It can be seen that there is a weak increase in use of *us* with *journey* over time ($r=0.3114$), while increase of *our* ($r=0.2806$) is close to that interpretation threshold. The corresponding plot shows that neither *our* nor *us* (the blue and orange lines respectively) are associated with *journey* initially, but become more correlated with *journey* over time, suggesting again that a particular use of metaphoric *journey* is adopted in this community. For the most-decreased variables *i* and *Negative*, there is little change in their relationship with *journey* over time, suggesting that use of

journey in the wider community outside cancer.net has a similar relationship with *i* and *Negative* as has been found in the cancer.net data.

Table 5-4 journey change in relationship with key variables over time

	min	max	mean	std	r	slope
journey_us	-0.0459	0.3498	0.0947	0.0708	0.3114	0.0004975
journey_our	-0.0515	0.327	0.1008	0.0751	0.2806	0.0004752
journey_i	-0.0137	0.346	0.0891	0.0561	0.0730	9.24e-05
journey_Negative	-0.0397	0.3333	0.0940	0.0559	0.0661	8.34e-05

Figure 5-9 correlation of journey with our, us, i, and Negative, over time



5.3.1.2 journey correlation over time with emotion variables

In the table below the correlation of *journey* with the eight emotion variables over time is listed by descending order of change. Although none of the results is significant according to the standard interpretation, there is a notable increase in prevalence of the positive emotion variables *Joy*, *Surprise*, *Anticipation*, *Positive*, and *Trust* over time, while none of the negative emotion variables changes notably.

Table 5-5 journey change in relationship with emotion variables over time

	min	max	mean	std	r	slope
journey_Joy	-0.1207	0.3224	0.1337	0.0613	0.2901	0.0004014
journey_Surprise	-0.1272	0.311	0.1168	0.058	0.2758	0.0003611
journey_Anticipation	-0.0295	0.2975	0.1363	0.0585	0.2587	0.0003416
journey_Positive	-0.0888	0.3117	0.1425	0.063	0.2585	0.0003677
journey_Trust	-0.084	0.3015	0.1345	0.0622	0.2373	0.0003331
journey_Fear	-0.0539	0.3163	0.0980	0.0557	0.0910	0.0001143
journey_Anger	-0.0518	0.3339	0.0956	0.0601	0.0844	0.0001146
journey_Negative	-0.0397	0.3333	0.094	0.0559	0.0661	8.34e-05
journey_Sadness	-0.0648	0.3774	0.0946	0.06	0.0398	5.39e-05
journey_Disgust	-0.0937	0.2879	0.0776	0.0542	0.0159	1.95e-05

The plots below further consider this pattern of change over time, the first plot for each of the emotion variables separately, and the second plot for the composite *Positive* and *Negative* emotion variables.

Figure 5-10 journey correlation with eight emotion variables over time

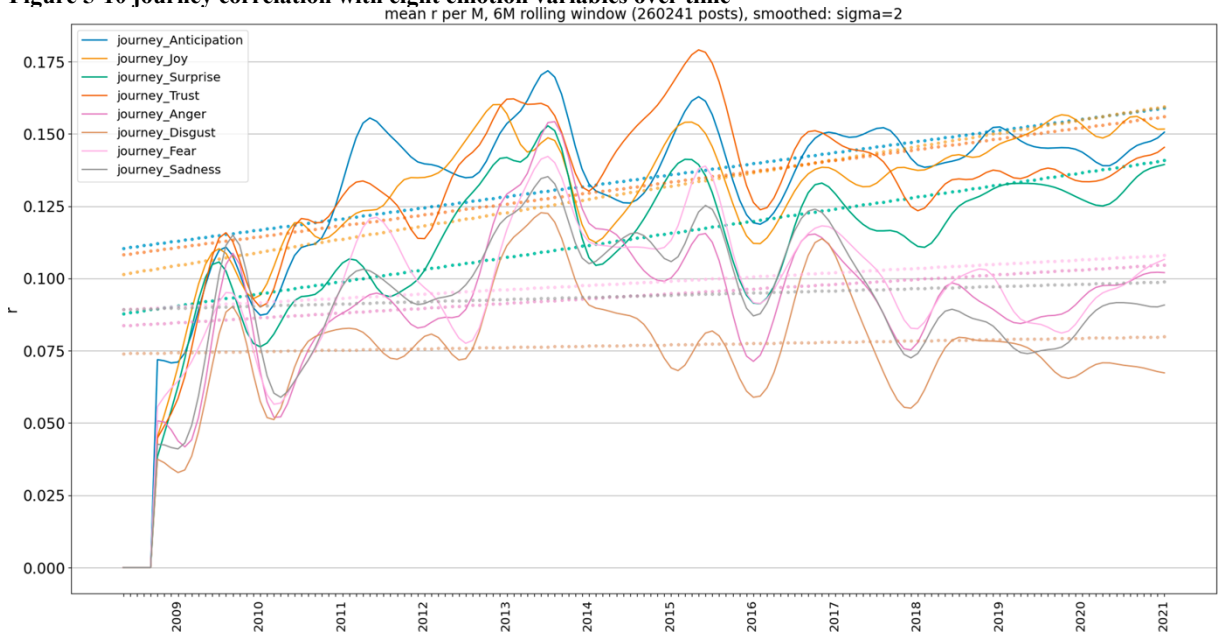
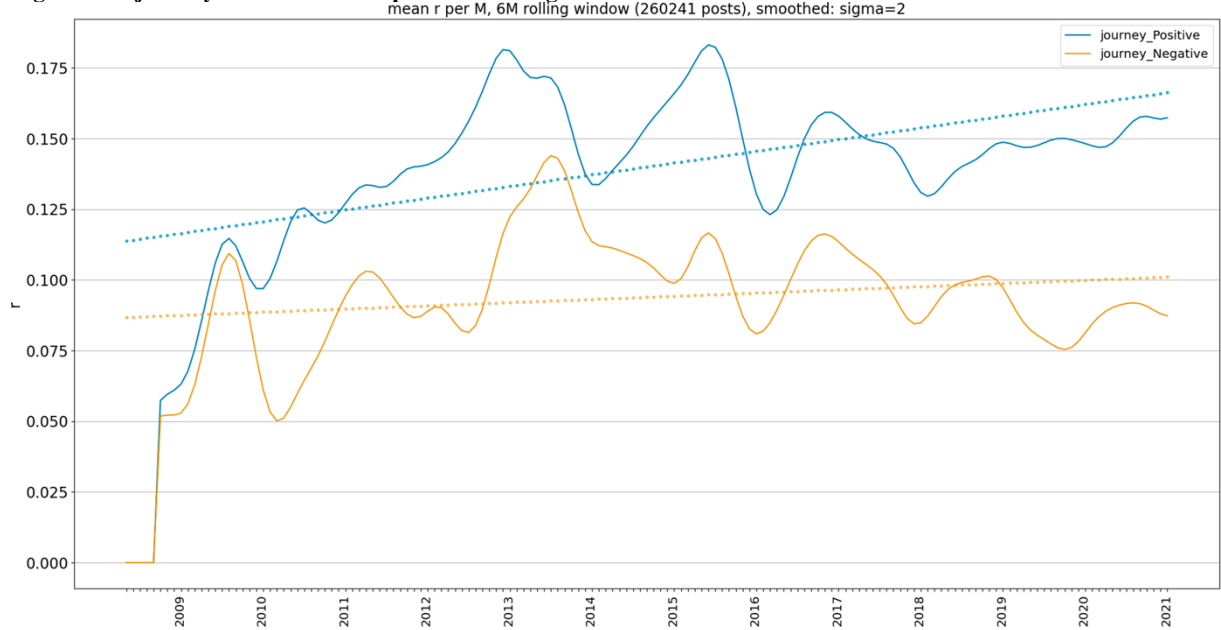


Figure 5-11 journey correlation with positive and negative emotion variables over time



The plots show that community use of positive and negative emotion variables in conjunction with *journey* track each other for the first 18 months, after which a separation occurs such that use of positive emotion variables in conjunction with *journey* increases in comparison to use of negative emotion variables. These results suggest that the *journey* metaphor is a vehicle for talking about emotions and becomes increasingly a way to talk about positive emotions over time on the forum. Along with the changes over time identified in connection with the key variables that are most changed in association with *journey*, this supports the interpretation in chapter two that the *journey* interpretative repertoire is concerned with the social and psychological work of supporting other participants emotionally in the context of the forum, and that it supports a focus on positive over negative emotions.

5.3.2 rollercoaster metaphor theme

5.3.2.1 rollercoaster correlation over time with most changed key variables

It was shown in chapter four that *fpp* and *we* are the most increased key variables, and *i* and *Trust* are the most decreased, in corpus F (posts in which *rollercoaster* is present) in

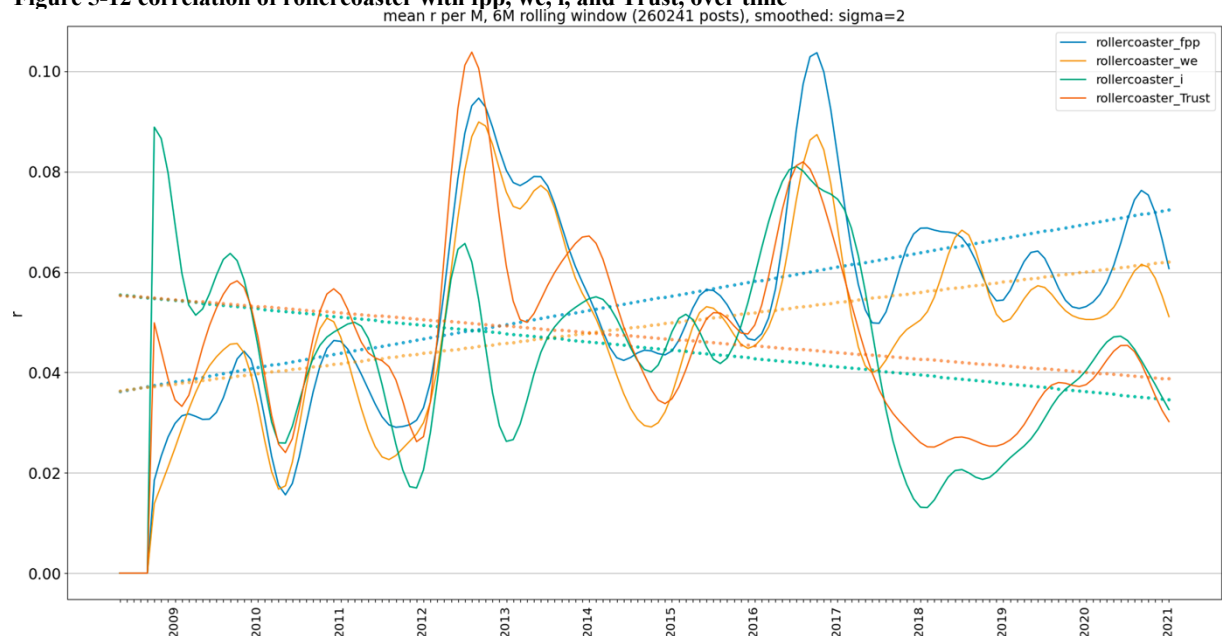
comparison to corpus R (posts in which *rollercoaster* is not present). The following table shows the minimum, maximum, mean, and standard deviation of r for the correlation of *rollercoaster* with each of these most-changed key variables by month, and their change over time (slope). It can be seen that although *fpp* shows the largest increase in use in conjunction with *rollercoaster* over time, and *i* shows the largest decrease, none of the key variables shows a significant change over time.

Table 5-6 change in use of rollercoaster in conjunction with key variables over time, all posts

	min	max	mean	std	r	slope
rollercoaster_fpp	-0.0633	0.2462	0.0545	0.0510	0.2033	0.0002338
rollercoaster_we	-0.0546	0.2184	0.0493	0.0500	0.1382	0.0001558
rollercoaster_i	-0.0295	0.3762	0.0461	0.0542	-0.1329	- 0.0001624
rollercoaster_Trust	-0.0309	0.2336	0.0486	0.0470	-0.1281	- 0.0001359

The plot below shows the pattern of change over time.

Figure 5-12 correlation of rollercoaster with fpp, we, i, and Trust, over time



5.3.2.2 rollercoaster correlation over time with emotion variables

It can be seen from the following table, which is ordered by absolute value of r , that there is no significant change in use of the emotion variables in conjunction with *rollercoaster*

over time. It is of note that use of all of the emotion variables decreases over time, suggesting that on cancer.net *rollercoaster* becomes increasingly less used to talk about specific emotions over time: it is more of a metaphoreme in itself representing sudden and potentially very notable changes in experience, without entailing the discussion of specific emotions. However, the change is not significant, and the plots show a varied picture over time, albeit with an overall downward trend.

Table 5-7 change in use of rollercoaster in conjunction with emotion variables over time

	min	max	mean	std	r	slope
rollercoaster_Sadness	-0.038	0.2519	0.0535	0.0563	-0.1610	-0.0002045
rollercoaster_Joy	-0.0226	0.2317	0.0484	0.0474	-0.1544	-0.0001652
rollercoaster_Negative	-0.0343	0.2514	0.0526	0.0537	-0.1489	-0.0001806
rollercoaster_Trust	-0.0309	0.2336	0.0486	0.047	-0.1281	-0.0001359
rollercoaster_Disgust	-0.0302	0.2892	0.0458	0.0508	-0.1250	-0.0001434
rollercoaster_Surprise	-0.0352	0.2863	0.0505	0.05	-0.1231	-0.0001389
rollercoaster_Anger	-0.0306	0.2178	0.0546	0.0535	-0.1141	-0.0001379
rollercoaster_Positive	-0.0271	0.3016	0.0511	0.0507	-0.1140	-0.0001305
rollercoaster_Anticipation	-0.0302	0.2279	0.0572	0.0466	-0.0969	-0.0001019
rollercoaster_Fear	-0.0383	0.2422	0.0497	0.0469	-0.0442	-4.68e-05

The plots below further consider the pattern of change over time, the first plot for each of the emotion variables separately, and the second plot for the composite *Positive* and *Negative* emotion variables. It can be seen that the emotion variables used in conjunction with *rollercoaster* largely track each other over time, and that there is an overall downward trend.

Figure 5-13 rollercoaster correlation with eight emotion variables over time



Figure 5-14 rollercoaster correlation with positive and negative emotion variables over time



5.3.3 fight metaphor theme

5.3.3.1 fight correlation over time with most changed key variables

It was shown in chapter four that the key linguistic variables *fpp* and *we* are the most increased, and *you* and *your* the most decreased, when corpus F (posts in which *fight* is present) is compared with corpus R (posts in which *fight* is not present). The first-person

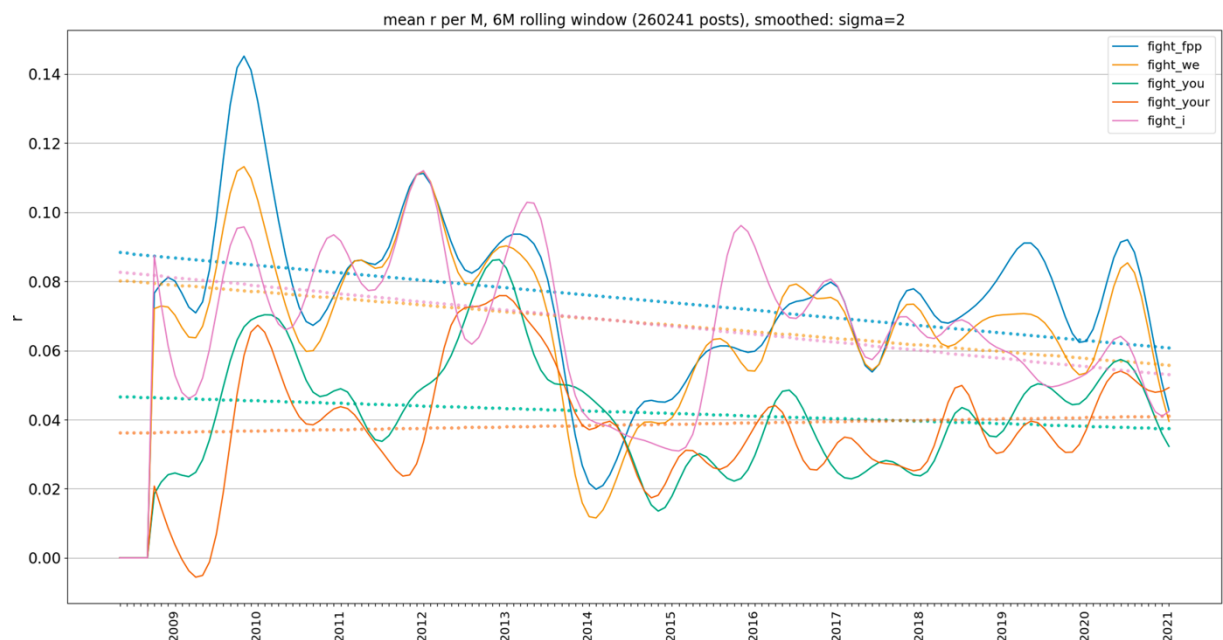
singular pronoun variable *i* was also included in the analysis in this section because of its primary importance in considering language style in connection with mental health. The following table, ordered by absolute value of *r*, shows the minimum, maximum, mean, and standard deviation of *r* for the correlation of *fight* with each of these variables, over time, by month. It can be seen that although the largest change over time is decreased use of *i* in conjunction with *fight*, none of the relationships over time is significant.

Table 5-8 change in use of fight in conjunction with key variables over time, all posts

	min	max	mean	std	r	slope
fight_i	-0.0528	0.2481	0.0679	0.053	-0.1633	-0.0001953
fight_fpp	-0.0282	0.3732	0.0746	0.0647	-0.1241	-0.0001814
fight_we	-0.036	0.3203	0.0681	0.0639	-0.1115	-0.0001608
fight_you	-0.0776	0.2063	0.0421	0.0451	-0.0601	-6.11e-05
fight_your	-0.0673	0.1823	0.0386	0.0439	0.0325	3.21e-05

The plot below shows the pattern of change over time.

Figure 5-15 change in use of fight in conjunction with key variables over time, all posts



5.3.3.2 fight correlation with emotion variables over time

The results table below shows that in contrast with the *journey* theme, the correlation of *fight* with each of the emotion variables decreases over time, although none of the changes is significant. It can be seen that use of *Disgust* words is the most decreased over time ($r=-0.2847$). As with the discussion of *rollercoaster* above, these results suggest that use of *fight* in this context comes to be associated with an interpretative repertoire that does not promote the finely grained discussion of emotion; on cancer.net this work is perhaps taken over over time by the *journey* metaphor.

Table 5-9 change of use of fight in conjunction with emotion variables over time

	min	max	mean	std	r	slope
fight_Disgust	-0.0042	0.2829	0.0802	0.0594	-0.2847	-0.0003817
fight_Negative	-0.0203	0.3578	0.0861	0.0652	-0.2811	-0.0004134
fight_Fear	-0.0213	0.3491	0.0872	0.0635	-0.2520	-0.0003609
fight_Sadness	-0.0222	0.3338	0.0854	0.0626	-0.2512	-0.0003546
fight_Anticipation	-0.0593	0.3824	0.0754	0.0604	-0.2498	-0.0003406
fight_Joy	-0.0423	0.3231	0.0720	0.0596	-0.2480	-0.0003334
fight_Positive	-0.0359	0.3241	0.0801	0.0595	-0.2165	-0.0002909
fight_Anger	-0.003	0.3056	0.0876	0.0581	-0.2142	-0.0002809
fight_Trust	-0.0367	0.2741	0.0746	0.0573	-0.2089	-0.0002700
fight_Surprise	-0.0683	0.3877	0.0767	0.061	-0.2077	-0.0002860

The plot below, which shows the pattern of change over time, demonstrates that use of each of the emotion variables in conjunction with *fight* largely track each other.

Figure 5-16 correlation of fight with eight emotion variables over time
 mean r per M, 6M rolling window (260241 posts), smoothed: sigma=2

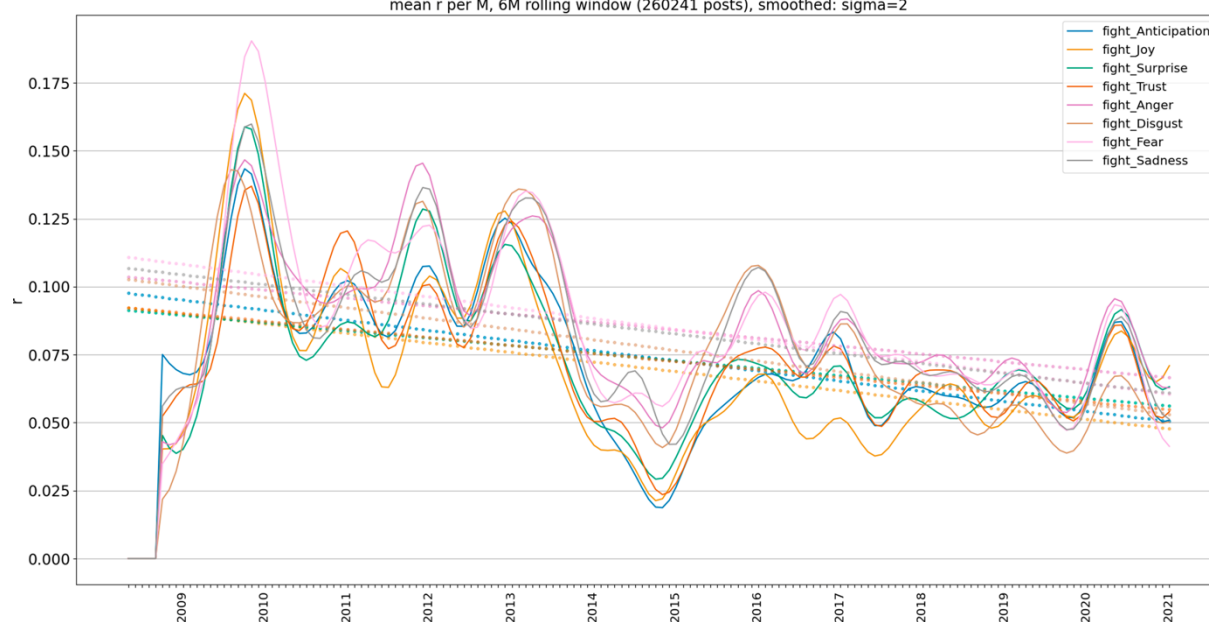
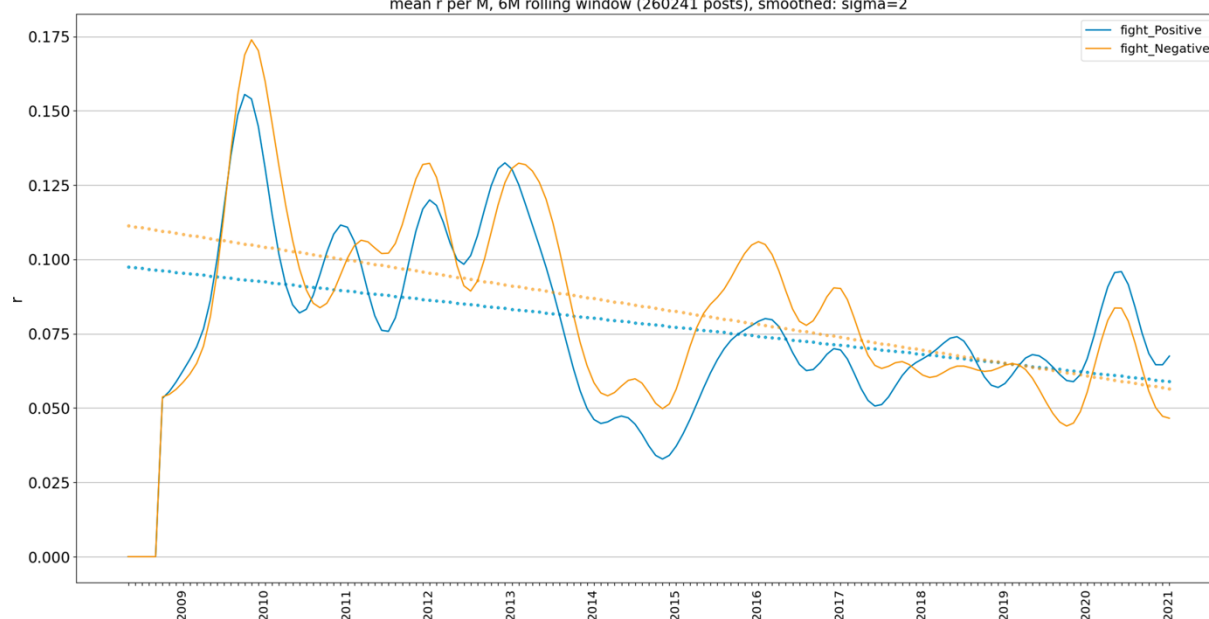


Figure 5-17 correlation of fight with Positive and Negative emotion variables over time
 mean r per M, 6M rolling window (260241 posts), smoothed: sigma=2



5.3.4 battle metaphor

5.3.4.1 battle correlation over time with most changed key variables

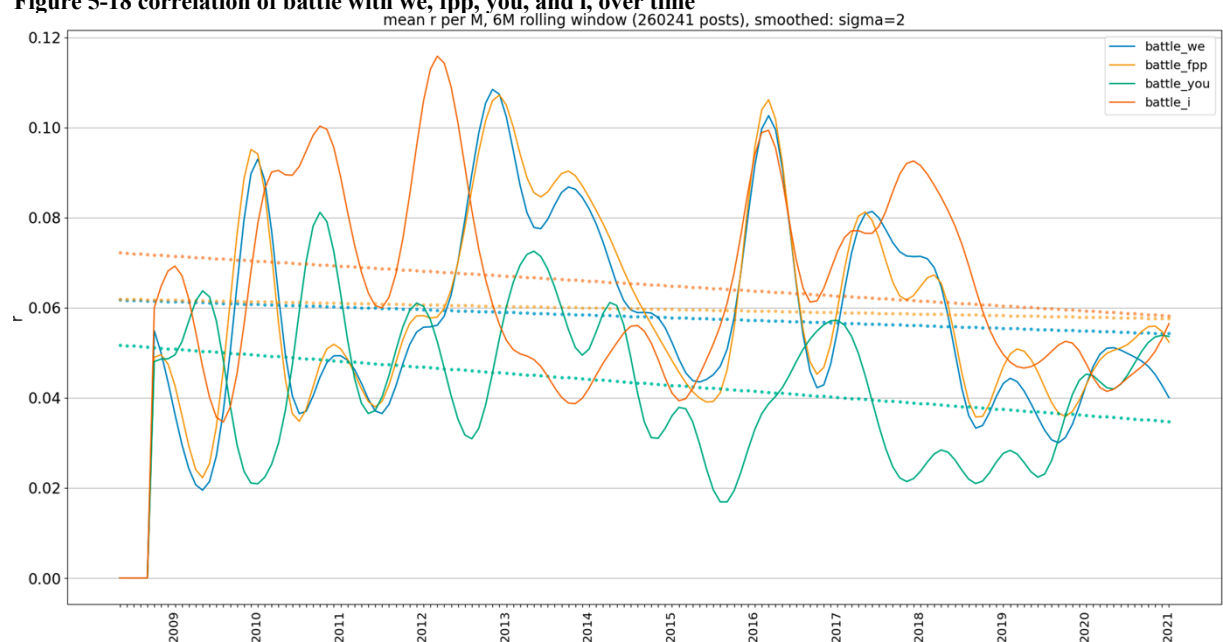
It was shown in chapter four that *we* and *fpp* are the most increased key variables, and *you* and *i* the most decreased, when corpus F (posts in which *battle* is present) is compared with corpus R (posts in which *battle* is not present). The following table,

ordered by absolute value of r , shows the minimum, maximum, mean, and standard deviation of r for the correlation of *battle* with each of these variables, over time, by month. It can be seen that although all of these key variables show a decrease over time in conjunction with use of *battle*, for none of them is this change notable. The fact that its associated work, identified through use of these key variables, does not change over time on cancer.net supports previous insights that *battle* is not adapted to do particular work in this community.

Table 5-10 change in use of battle in conjunction with key variables over time, all posts

	min	max	mean	std	r	slope
battle_you	-0.0297	0.2792	0.0437	0.0487	-0.0916	-0.0001007
battle_i	-0.053	0.2342	0.0649	0.0536	-0.0348	-4.21e-05
battle_fpp	-0.031	0.2984	0.0603	0.0557	-0.0281	-3.53e-05
battle_we	-0.0405	0.3011	0.0577	0.0563	-0.0204	-2.59e-05

Figure 5-18 correlation of battle with we, fpp, you, and i, over time



5.3.4.2 correlation of battle with emotion variables

The results in the table below show that, as with the *rollercoaster* and *fight* themes, the correlation of *battle* with each of the emotion variables decreases over time, although none of the results are significant. It can be seen from the table, which is ordered by

absolute value of r , that use of *Anticipation* words is the most decreased over time ($r=-0.1808$). As with the discussion of *rollercoaster* and *fight* above, this suggests that use of *battle* in this context comes to be associated with an interpretative repertoire that does not promote the finely grained discussion of emotion, and that on cancer.net this work is perhaps taken over over time by the *journey* metaphor.

Table 5-11 change in use of battle in conjunction with emotion variables over time

	min	max	mean	std	r	slope
battle_Anticipation	-0.0381	0.3166	0.0766	0.059	-0.1808	-0.0002409
battle_Joy	-0.039	0.2403	0.0691	0.0528	-0.1623	-0.0001935
battle_Surprise	-0.0255	0.2429	0.0755	0.0554	-0.1515	-0.0001893
battle_Negative	-0.0582	0.3974	0.0846	0.0665	-0.1338	-0.0002008
battle_Fear	-0.0256	0.3715	0.0869	0.0656	-0.1118	-0.0001655
battle_Positive	-0.0292	0.2488	0.0777	0.0587	-0.0948	-0.0001255
battle_Trust	-0.0353	0.2955	0.0737	0.0604	-0.0906	-0.0001235
battle_Sadness	-0.0362	0.3754	0.0857	0.0633	-0.0842	-0.0001202
battle_Anger	-0.0493	0.3214	0.0926	0.0626	-0.0751	-0.0001061
battle_Disgust	-0.0474	0.2662	0.0821	0.0589	-0.0737	-9.81e-05

The plot below demonstrates that, as for the *rollercoaster* and *fight* themes discussed in previous sections, use of each of the emotion variables in conjunction with *battle* largely track each other over time.

Figure 5-19 correlation of battle with eight emotion variables over time

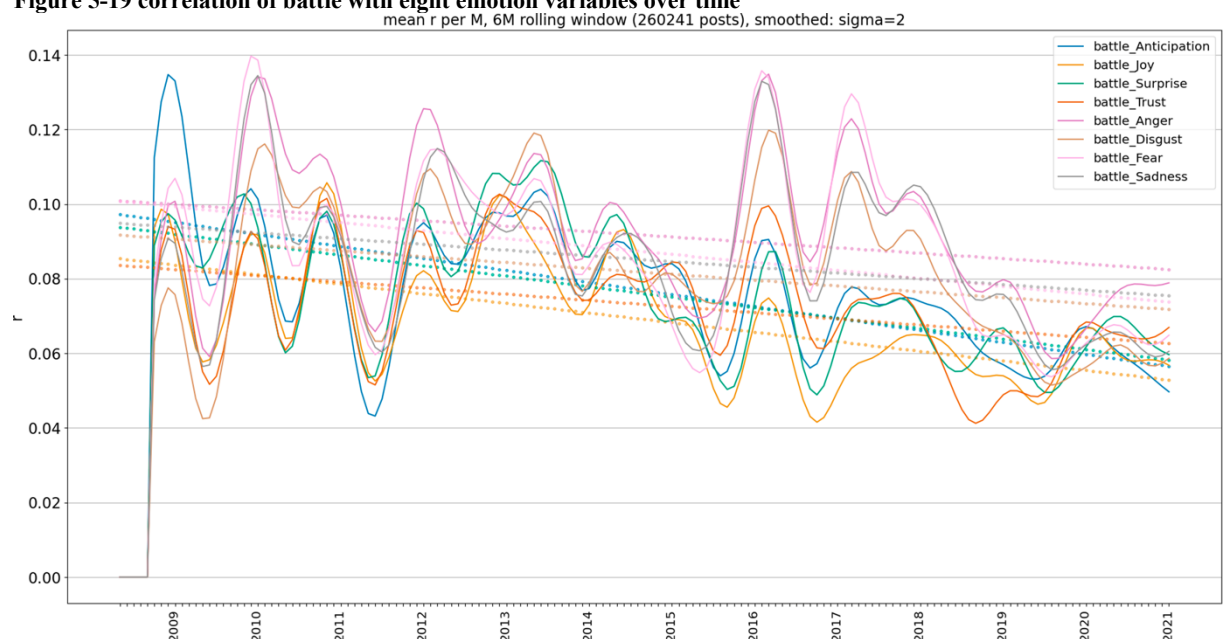
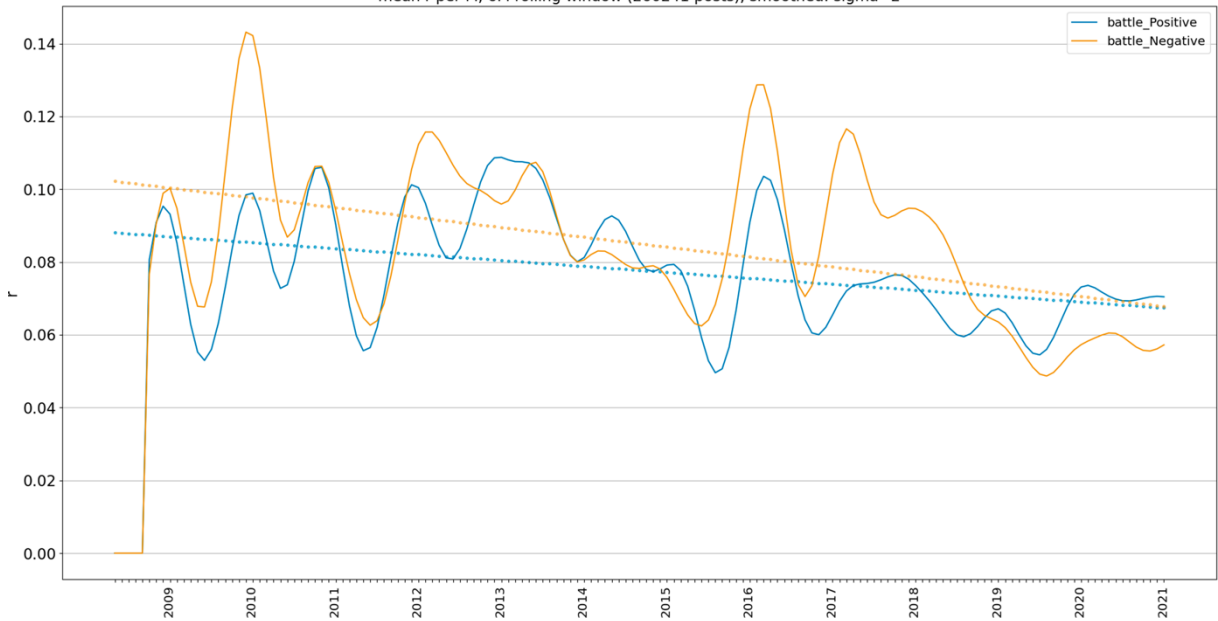


Figure 5-20 correlation of battle with Positive and Negative over time
 mean r per M, 6M rolling window (260241 posts), smoothed: sigma=2



5.3.5 Section summary: correlation of metaphor themes with key variables over time

The correlation of each separate metaphor theme over time with key variables relating to state of mind, and emotion variables, provides insight into their changing role as a vehicle for social and psychological work in this community. Use of *rollercoaster*, *fight*, and *battle*, were found to be associated with a decreased use of emotion variables over time, while for *journey*, an association with use of all emotional variables increases over time, however this relationship is notable for the positive emotion variables only. This suggests that the *journey* metaphor comes to be associated with, and to dominate, emotion work on cancer.net, but most specifically with emotion work related to positive emotions, such that negative emotion discourse may be suppressed.

5.4 Use of key variables for comparative participant groups over time

To investigate the effect of participation in the cancer.net community on language style, in this section consideration is given to the change in density of key variables for participants in the different stratified groups described in the table below. For the groups

of more established participants (cancer.net.20 and cancer.net.200) it should be noted that their data also includes the posts those participants made when they were new to the forum, i.e. each group contains all the data for the selected participants, from their first post to their last post.

Table 5-12 comparative participant groups

posts greater than	posts less than	referred to as	plot line
0	20	cancer.net.LT20	green
19		cancer.net.20	yellow
199		cancer.net.200	blue

The first analysis in the current section focuses on change in use of key pronouns characterising particular state of mind in relation to language style over time, *our*, *we*, *i*, *me*, and *my*, for the comparative participant groups. Following this an analysis is made of the change in use of the metaphor themes over time. These analyses provide insights into which key variables, and which metaphor themes, come to be more used with increased participation on cancer.net

5.4.1 Change in density of key pronouns for different participant groups over time

In this analysis, in order to consider the general effect of participation on cancer.net over time separately to use of the metaphor themes, comparative data was selected based on the number of posts made by participants. In order to make the data easier to compare between groups, the individual tables containing the results for each participant group are listed first, followed by the related plots showing the change in density over time of the key pronouns for each group.

Table 5-13 change in density over time for key pronouns, for participants who have made fewer than 20 posts
cancer.net **min** **max** **mean** **std** **r** **slope**
.LT20

our	0.0002	0.0021	0.0007	0.0003	-0.2048	-1.3e-06
we	0.0017	0.0068	0.0039	0.001	-0.2848	-6.3e-06
i	0.0334	0.0479	0.0412	0.0031	0.4388	3.11e-05
me	0.0038	0.0091	0.0058	0.0009	-0.1743	-3.4e-06
my	0.0111	0.0184	0.0148	0.0014	0.3737	1.15e-05

Table 5-14 change in density over time for key pronouns, for participants who have made 20 or more posts
cancer.net **min** **max** **mean** **std** **r** **slope**
.20

our	0.0001	0.0023	0.001	0.0004	0.2072	1.6e-06
we	0.0014	0.0063	0.0039	0.0008	-0.0923	-1.7e-06
i	0.024	0.0406	0.0313	0.0036	-0.1674	-1.36e-05
me	0.0025	0.0067	0.0045	0.0008	-0.2115	-3.6e-06
my	0.0068	0.0136	0.0097	0.0011	-0.1420	-3.4e-06

Table 5-15 change in density over time for key pronouns, for participants who have made 200 or more posts
cancer.net **min** **max** **mean** **std** **r** **slope**
.200

our	0.0	0.0028	0.0011	0.0005	0.3329	3.8e-06
we	0.0003	0.0072	0.0039	0.0011	0.1594	4.1e-06
i	0.0177	0.0443	0.0267	0.0041	-0.2710	-2.5e-05
me	0.0009	0.006	0.0037	0.0009	-0.0364	-8e-07
my	0.004	0.0146	0.0081	0.0014	-0.0191	-6e-07

In terms of change relating to the comparative participant groups, it can be seen from the tables above that the density of use of both of the first-person plural pronouns (*our*, and *we*) increases with increased participation on the forum, while density of use of *i* decreases. The pronoun that increases most with increased participation is the first-person plural pronoun *our* ($r=-0.2048$, $r=0.2072$, $r=0.3329$), while the pronoun that decreases most is the first-person singular pronoun *i* ($r=0.4388$, $r=-0.1674$, $r=-0.271$).

Comparative plots for each of the most changed pronouns *our* and *i* for each participant group over time are included below. The line for cancer.net.LT20 is green, for cancer.net.20 it is yellow, and for cancer.net.200 it is blue. It can be seen from these

plots, as was shown in the tables above, that density of use of *our* (first plot) tends to be higher for more established participant groups, while density of *i* (second plot) tends to be lower for more established participants groups. In addition, the flat (blue) line relating to use of *our* at the start of the first plot below, which is the line representing the most established participants who have made more than 200 posts, shows that use of this community referencing pronoun was not typical in the first year of the forum. This effect is not apparent in the subsequent plot, relating to the first-person singular pronoun *i*, since use of *i* is not dependent on a sense of community: it is used throughout the cancer.net data.

Figure 5-21 change in density of *our* over time for different participant groups

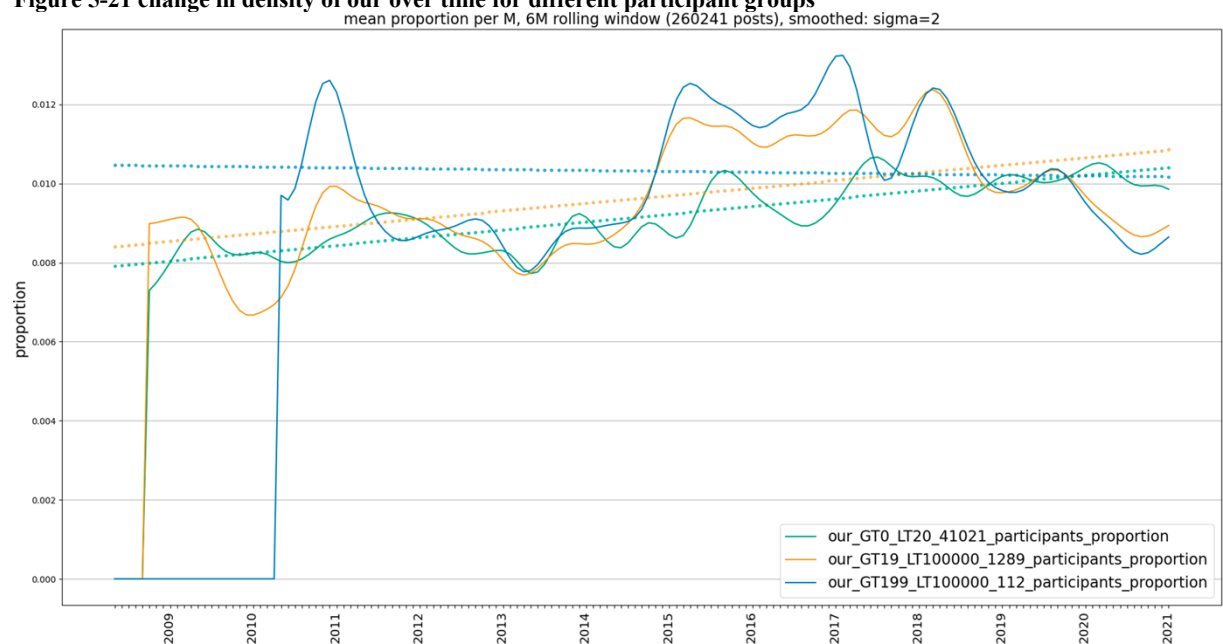
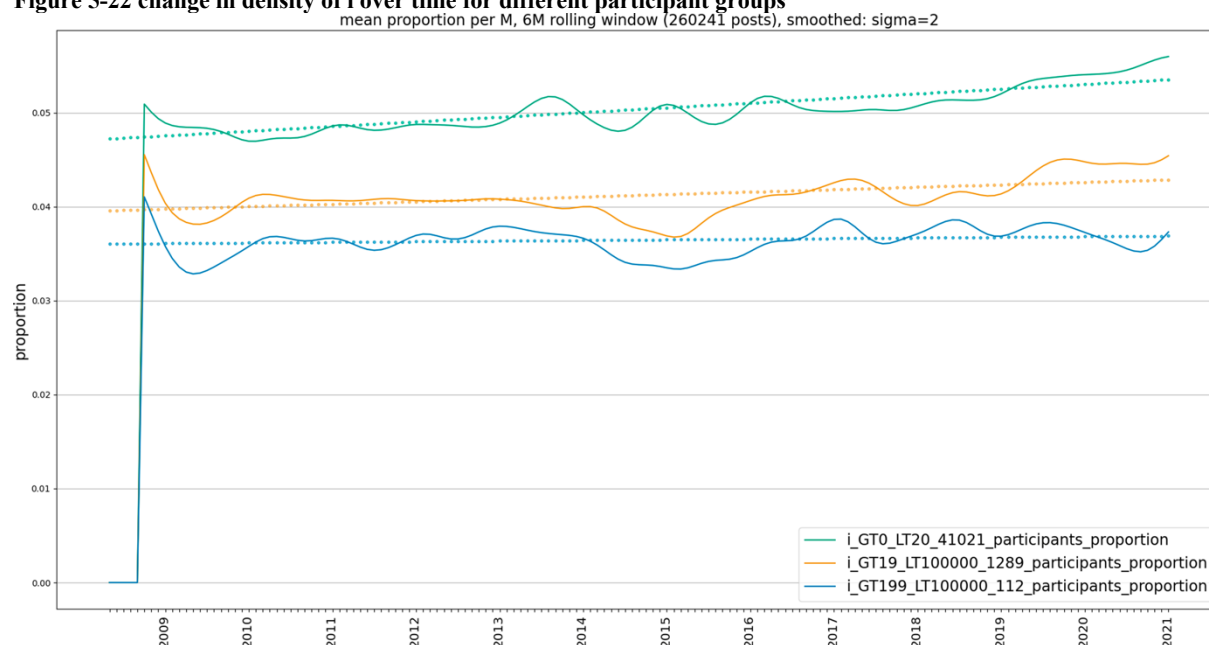


Figure 5-22 change in density of i over time for different participant groups



5.4.2 Change in density of metaphor themes for different participant groups over time

Following on from consideration of the change in density of key pronouns in the previous section, in this section the change in density of use over time of the four metaphor themes for each participant group is reported in the tables below, to provide further insight into their identity as community metaphors, with such metaphor expected to become more densely used by more established participants.

Table 5-16 change in density over time for metaphor themes, for participants who have made fewer than 20 posts

cancer.net.LT20	min	max	mean	std	r	slope
journey	0.0	0.0588	0.0183	0.0091	0.0909	1.87e-05
rollercoaster	0.0	0.0588	0.0062	0.0069	-0.1273	-1.99e-05
fight	0.0	0.1176	0.0115	0.0118	-0.3095	-8.23e-05
battle	0.0	0.0447	0.0116	0.008	-0.2734	-4.91e-05

Table 5-17 change in density over time for metaphor themes, for participants who have made 20 or more posts

cancer.net.20	min	max	mean	std	r	slope
journey	0.0	0.0845	0.0359	0.0159	0.1886	6.75e-05
rollercoaster	0.0	0.0278	0.0064	0.0047	0.1406	1.51e-05
fight	0.0	0.0312	0.0092	0.0072	-0.5246	-8.53e-05
battle	0.0	0.0482	0.0088	0.0073	-0.3922	-6.47e-05

Table 5-18 change in density over time for metaphor themes, for participants who have made 200 or more posts cancer.net.200

	min	max	mean	std	r	slope
journey	0.0	0.1062	0.0391	0.0215	0.2358	0.0001146
rollercoaster	0.0	0.0291	0.0056	0.0057	0.2406	3.12e-05
fight	0.0	0.0556	0.0089	0.0094	-0.3641	-7.71e-05
battle	0.0	0.0556	0.0074	0.0099	-0.3200	-7.14e-05

It can be seen from the tables above that the density of use of both *journey* ($r=0.0909$, $r=0.1886$, $r=0.2358$) and *rollercoaster* ($r=-0.1273$, $r=0.1406$, $r=0.2406$) increases over time with more participation on the forum, while the density of use of *fight* ($r=-0.3095$, $r=-0.5246$, $r=-0.3641$) and *battle* ($r=-0.2734$, $r=-0.3922$, $r=-0.32$) decreases over time for each participant group.

5.4.3 Section summary: use of key variables for comparative participant groups over time

A comparison of the density of key variables for different participant groups based on number of posts made on cancer.net, representing level of establishment within the cancer.net community, showed that use of the first-person plural pronouns *our* and *we* increases with increased participation on the forum, and use of the first-person singular pronoun *i* decreases, while the pattern is more mixed for the first-person singular pronouns *me* and *my*. This is a language style that has been associated with better mental health, suggesting that participation in the cancer.net community is beneficial to mental health. It shows that the subject position of participants evolves with increased participation to reflect their sense of membership in the community. In addition, it was shown that use of the *journey* and *rollercoaster* metaphor themes increases with increased participation on cancer.net, while use of *fight* and *battle* decreases over time for each participant group. This suggests that in terms of metaphor, membership in this community entails adoption of the *journey* and *rollercoaster* themes, and suppression of the *fight* and *battle* themes.

5.5 Diachronic analysis discussion

In the diachronic analysis carried out in this chapter it was shown that use of the *journey* metaphor theme becomes more dominant over time on cancer.net and use of the *fight* and *battle* themes in comparison decreases significantly over time. In addition, use of the *journey* and *rollercoaster* themes for individual participants increases in general with increased participation on the forum, while use of *fight* and *battle* decreases over time for all participant groups. It is also notable that density of use of all the metaphor themes decreases with an increase in new participants to the forum and increases in line with the levelling off of new participants to the forum. This suggests that regardless of whether they have been adopted and adapted to do particular work in the community, in cancer.net potentially more deliberate metaphor such as those represented by the identified metaphor themes tend not to be used by participants who are new to the community, who are likely to have more recently begun their experience of living with cancer, and to be more focused on the literal detail of their own current lived experience.

Use of *rollercoaster*, *fight*, or *battle*, is associated with a decreased use of emotion variables over time, while for *journey* an association with use of all emotion variables increases over time, however this relationship is notable for positive emotion variables only. This suggests that the *journey* metaphor comes to be associated with, and dominant with regard to, emotion work on cancer.net, but most specifically with emotion work related to positive emotions.

A comparison of the density of key variables for different participant groups based on number of posts made on cancer.net, representing level of establishment within the cancer.net community, showed that use of the first-person plural pronouns *our* and *we* increases with increased participation on the forum, while use of the first-person singular

pronoun *i* decreases, suggesting an increased feeling of belonging in this community, which moves the focus away from the self. This is a language style that has been associated with better mental health, suggesting that participation in the cancer.net community is beneficial to mental health. This may relate to metaphor, in that use of the *journey* and *rollercoaster* themes for individual participants also increases in general with increased participation on the forum, while use of *fight* and *battle* decreases over time for each participant group. This supports the view that community metaphors and the interpretative repertoires, the ethos and rhetoric, associated with them, are an important and powerful factor in the identity of the cancer.net community.

5.6 Cancer.net discussion

In chapter one of this thesis the primary prediction was stated as follows: where metaphor is used to characterise a concept, the surrounding language will be of a style that has been found to be associated with better mental health. This has been shown to be true for cancer.net in terms of pronoun use associated with better mental health. It has been shown that where any of the identified metaphor themes *journey*, *rollercoaster*, *fight*, or *battle*, are present in a post, use of the first-person singular pronoun *i* is significantly decreased, and use of the first-person plural pronoun *we* and the composite first-person plural pronoun *fpp* are significantly increased. This effect was found to be strongest for the *journey* theme, which in addition co-occurs with significantly increased use of the first-person plural pronouns *our*, and *us*, which are the two most increased variables associated with the *journey* theme. In addition, all of the metaphor themes were found to be associated with use of a wider range of pronouns, another pronoun style that has been associated with better mental health. The *journey* theme was also found to be associated with increased use of positive emotion variables, and decreased use of negative emotion variables. So, while all of the themes were found to be associated with elements of

language style that have been associated with better mental health, use of the *journey* theme was found to be associated with a more extensive language style representative of better mental health than were the other themes.

A separate analysis considered each metaphor theme in isolation from the other themes. The notable fact that use of the first-person plural pronoun *our* is not significantly increased when *journey* is used in isolation from other themes, when it is the most increased variable when all instances of *journey* were considered, suggests that *our* is used with *journey* specifically in conjunction with other themes, perhaps in the context of an ideological dilemma. The pronoun *our*, then, gives an insight into the work of a metaphor-related ideological dilemma of this community, which could specifically relate to negotiation of a community stance towards the experience of living with cancer.

Use of the key variable *i* was found to be decreased only for the *journey* and *rollercoaster* themes used in isolation, and not for the *fight* and *battle* themes. In terms of subject positions this suggests that participants are using the *fight* and *battle* themes in isolation to communicate their individual experience. This is consistent with the findings in chapter three that *battle* is more used by newer than by more established participants, while *fight* is equally used by both groups. The significant increase in use of the first-person plural pronoun *we* in conjunction with each theme shows that where any metaphor theme is used, either in isolation or combined with other metaphor themes, there is a sense of membership, and a perspective that looks beyond the self.

A more detailed consideration of the work of the metaphor themes comes from the topic modelling investigation in chapter two, in which comparisons were made between topics instantiated as nouns, verbs, and adjectives, and in which the constituent terms were included at different, comparative, density levels. In the condition in which only nouns

were included in the topic modelling analysis and were only included where they are present in 10% or fewer of posts, a single topic was found to contain the potentially metaphoric terms *journey*, *step*, *rollercoaster*, *road*, and *fight*. This topic appears to be associated with warmth and familiarity in conjunction with the forum, since alongside technical terms related to the forum, such as *post*, *thread*, *word*, *message*, *reply*, *quote*, and *comment*, it includes the emotional terms *love*, *hug*, *hope*, *thought*, *strength*, *good*, and *inspiration*, and the familiarly-phrased person references *lady*, *lass*, *folk*, and *buddy*. When more prevalent nouns were included in the topic output, that are present in 20% or fewer of documents, the topic in which *journey* and *rollercoaster* are situated appears to support discussion of emotion terms. When more common nouns, verbs, and adjectives were included, use of the *journey* metaphor appears to support use of emotion terms, personal terms, and the wider-focus terms *world*, *god*, and *life*, suggesting that the interpretative repertoire containing *journey* is used to talk about life as a whole, and about concepts that are wider than a single life. In addition, the high coherence of the topics in each investigative condition supports the view of cancer.net as a community in that it represents the coherence of the common concerns and focus of this large and diverse forum population.

In chapter three, in which the actual metaphoricity of potentially metaphoric terms was confirmed, an analysis of signalled metaphor in conjunction with the dominant metaphor themes of this community showed that there is a significant difference in the density of signalled metaphor between the metaphoric conditions ($p < 0.001$), with the density of every metaphor signalling pattern lowest for the condition in which none of the metaphor themes are present. It is of particular note that the effect size is very high for copular similes ($\chi^2 = 1,943$, $p < 0.001$), which were found to be predominantly very creative.

The biggest effect size ($\chi^2=3,024$, $p<0.001$) was found for the *a sort of* and *a kind of* constructions that are specifically excluded from MIPVU. This suggests that such vague language plays a particular role in conjunction with metaphor, which is itself slippery and malleable such that it may be adapted to suit evolving purposes. The use of vague language constructions such as *a sort of* and *a kind of* has been associated with the discussion of sensitive topics, such that as the level of sensitivity of a topic increases, so the level of vagueness in talk-in-interactions also increases (Zhang, 2013). The increased use of these vague terms in conjunction with community metaphor may indicate that such potentially more deliberate, or more active, metaphor supports the discussion of sensitive topics; it may also be associated with the evolution of the work of the metaphor themes over time.

Another characteristic of the work that is being done by the metaphor themes in this context was identified in chapter three, when graph-based ranking and an investigation into the prevalence of noun phrases used by more established in comparison to less established participants identified the phrase *cancer journey* as a metaphoreme, which was also found to be used in conjunction with the *rollercoaster* and *battle* themes. Fewer time references were associated with the *journey* and *rollercoaster* themes; for the *journey* theme these were *year journey* and *life journey*, while for *rollercoaster* the phrase *life rollercoaster* was used. For the *fight* and *battle* themes, specific time references are more common, and they are present in over 50% of *battle* phrases, and in 70% of the *battle* phrases of more recent participants. Time references for *battle* include *day*, *week*, *month*, *half year*, and *year*, but for neither *fight* nor *battle* do time references include *life*.

The *cancer journey* metaphoreme, then, may be said to frame the experience of cancer as a longer-term endeavour, such that cancer is typically not the sudden, catastrophic cliff

edge, that it might appear to be when first encountered, and neither is it the shorter term *cancer battle* of a *day, week, month, half-year*, and also *year*. The *cancer journey* is consistent with the difficulties that are inherent in any life journey and does not entail the end of life in the short term. This metaphoric framing of the experience of cancer is part of the shared knowledge and ethos on cancer.net that is adopted with exposure to, and participation on, the forum. For example, the *journey* and *rollercoaster* themes were found to be more prevalent in data for the more established cancer.net participants who have made 20 or more posts, while *battle* is more prevalent in data for participants who have made fewer than 20 posts, and *fight* is equally prevalent for both sets of participants.

The second aim of this thesis is to consider the role of metaphor in the formation and evolution of the normative stance of a community over time, by considering change in density of metaphor and other key variables in the data as a whole, and for comparative participant groups based on number of posts made within the forum.

Use of the *rollercoaster*, *fight*, or *battle*, themes was found to be associated with a decreased use of emotion variables over time, in comparison for *journey* use of all the emotional variables increases over time, however this relationship is notable for the positive emotion variables only. This suggests that the *journey* metaphor comes to be associated with, and dominant with regard to, emotion work on cancer.net, but most specifically with emotion work related to positive emotions, such that it may be part of the ethos of this community to suppress negative emotion discourse.

A comparison of the density of key pronouns for different participant groups based on number of posts made on cancer.net, representing level of establishment within the cancer.net community, showed that use of the first-person plural pronouns *our* and *we* increases with increased participation on the forum, suggesting an increasing feeling of

belonging in this community, while use of the first-person singular pronoun *i* decreases. This is a language style that has been associated with better mental health, suggesting that participation in the cancer.net community is beneficial to mental health. This is associated with metaphor use in that use of the *journey* and *rollercoaster* themes also increases with increased participation on the forum, while use of *fight* and *battle* decreases over time for each participant group. It is of particular note that use of the community referencing pronoun *our* was not typical in the first year of the forum, while other key pronouns analysed, such as *i*, are used by participants throughout their participation on the forum.

In terms of the third aim of the thesis, then, the established corpus linguistic techniques and other exploratory techniques from data science used in the current investigation have been productive in finding metaphor, and in establishing what the social and psychological work of metaphor is in the computer-mediated discourse of this virtual community.

In summary, although *journey* is not the most prevalent theme at the start of the forum, it comes to be the dominant metaphor theme on cancer.net. The social and psychological work of the *journey* metaphor on cancer.net includes: support for better mental health through creating a feeling of belonging and support; discussion of emotions, particularly positive emotions; relative suppression of the discussion of negative emotions; relative suppression of the *fight* and *battle* metaphor themes; and framing the experience of living with cancer in a wider timeframe and context.

In the next chapter some of the quantitative and qualitative techniques explored to evaluate the social and psychological work of metaphor on cancer.net are applied to a pro-choice suicide forum. In the subsequent, final, chapter of this thesis consideration is given to the differences between the work that metaphor appears to be doing on the

cancer.net forum, which is focused on discussion of an existentially threatening physical illness, and the work that metaphor appears to be doing on the suicide.net forum, which is focused on existentially threatening attitudes and mental concerns.

6 SUICIDE.NET

6.1 Introduction

In chapters two to five of this thesis, the primary research question, whether where metaphor is used to characterise a concept the surrounding language is of a style that has been found to be associated with better mental health, has been answered in the context of cancer.net. The initial prediction that this would be the case, as represented in the first example in chapter one (section 1.1.2), has been found to be true in the cancer.net context. This has been related to and supported by the second novel contribution of the thesis, to consider the role of metaphor in the formation and evolution of a community over time. The third novel contribution of this thesis, the application of new techniques from data science and natural language processing alongside more established corpus linguistic techniques (section 1.1.1.1) in terms of (i) finding metaphor in corpora; (ii) analysing language style; and (iii) diachronic analysis; has been fulfilled in the consideration of the primary and secondary research questions applied to the cancer.net corpus.

To consider the more general applicability of the methods used, and to understand more about the generalizability of the results from cancer.net, including comparison of results between a corpus relating to predominantly physical health concerns (cancer.net) and a corpus relating to predominantly mental health concerns (suicide.net), in chapter six methods applied in chapters two to five to the cancer.net corpus are applied to the suicide.net corpus. However, in this second application of methods there is less discussion about the actual methods, resulting in a more condensed analysis that spans a single chapter. The 'pro-choice' suicide forum suicide.net was selected for this purpose because of the high-stakes life or death outcomes attributed to its existence, or cessation,

from very different cultural perspectives. As is discussed in more detail in a subsequent section, while the internet "can be an invaluable resource for individuals experiencing self-harm and suicidal feelings" (Samaritans, 2020, p. 3), it may also exacerbate such behaviours (Samaritans, 2020, 2021), such that this is a pressing health care issue requiring further investigation and understanding (Bell, 2007; Kirmayer, Raikhel and Rahimi, 2013; Marchant *et al.*, 2017). In the UK suicide is the leading cause of death of men aged 20-49, and of women aged 20-34, while globally it is the second leading cause of death in those aged 15-29, after road traffic accidents (ONS, 2019).

Suicide is typically a taboo subject, and consequently is often misconstrued. The folk psychology around suicide can be dismissive and unhelpful, for example "those who talk about suicide are not at risk of suicide"; "suicidal behaviour is motivated by attention-seeking"; "suicide cannot be prevented"; "thinking about suicide is rare" (O'Connor, 2021, p. 49). As a powerful social and psychological mechanism metaphor may facilitate communication around suicide, and improve understanding; for example, the following evocative extended metaphor about suicide and suicidal ideation by the novelist David Foster Wallace, who killed himself aged 46:

The so-called psychotically depressed person who tries to kill herself doesn't do so out of quote hopelessness or any abstract conviction that life's assets and debits do not square. And surely not because death seems suddenly appealing. The person in whom Its invisible agony reaches a certain unendurable level will kill herself the same way a trapped person will eventually jump from the window of a burning high-rise. Make no mistake about people who leap from burning windows. Their terror of falling from a great height is still just as great as it would be for you or me standing speculatively at the same window just checking out the view; i.e. the fear of falling remains a constant. The variable here is the other terror, the fire's flames: when the flames get close enough, falling to death becomes the slightly less terrible of two terrors. It's not desiring the fall; it's terror of the flames. And yet nobody

down on the sidewalk, looking up and yelling Don't! and Hang on!, can understand the jump. Not really. You'd have to have personally been trapped and felt flames to really understand a terror way beyond falling. (Wallace, 2014)

6.1.1 Suicide.net background

Suicide.net started in November 1990 as the alt UseNet newsgroup alt.suicide.holiday, which was set up as a place to discuss the relationship between suicide rates and holidays. In 2013 it moved to the Reddit discussion website as the r/sanctionedsuicide subreddit, but was banned on 14th March 2018 for breaking Reddit guidelines, specifically for encouraging violence. The closure was received with sadness and dismay, for example: "this feels like vandalism, like burning someone's letters. And demolishing the headstones in a graveyard" (r/PsychiatricFreedom - sanctioned suicide banned, 2018).

Suicide.net was re-started on 18th March 2018 at the independent web address sanctioned-suicide.com, which is where the data used in the current study was collected in June 2020. The sanctioned suicide identity is a reference to the idea that certain forms of suicide are culturally sanctioned, for example Japanese seppuku, a ritual form of suicide used to restore honour; euthanasia; and terrorist martyrdom. These culturally-sanctioned forms of suicide are not regarded as actual suicide in that they are not equatable with depression and other mental illness, and therefore do not carry the shame typically associated with suicide (Pierre, 2015).

Following further controversy, allegedly because of the presence of minors on the site, the sanctioned-suicide.com domain name was banned by its registrar in January 2021. The site was moved in January 2021 to the new url sanctioned-suicide.org. In a softening of its original stance it displays prominently on every page: a warning that the site is not for

minors; the US Samaritans telephone number; and a link to a page of other suicide resources which includes, for example, a link to the UK Samaritans website.

The public persona of suicide.net is an apt metaphor for the personal experience of its participants, who feel misunderstood, persecuted for being different, and at odds with hegemonic culture: "normies will do everything to get this site taken down". For example, "[i] correlate pretty hard with asperger's descriptions of feeling like an alien or freak in this world, like everyone's playing a game that they received instructions for, and we never got our manual". Recent BBC news articles describe the forum as having played a part in the UK suicides of a 23 year old woman with autism (BBC News, 2020), and a 25 year old man with autism (BBC News, 2021). In the latter article, his mother specifically refers to the suicide.net age warning and advertising of the Samaritans helpline as insufficient for vulnerable people such as her son. She sees a potential solution in the removal of the stigma of talking about suicide:

People are fearful of saying the word suicide, it's hidden, so people who are suffering with mental health and thinking of this, it's to get awareness to say, it's okay to feel this way, so we can deal with that and get help (BBC News, 2021)

Individuals diagnosed as autistic are around three times more likely to have suicidal ideation than are the non-autistic population (Cassidy and Rodgers, 2017; Kølves *et al.*, 2021). This has been related to the inherent autism issue that thoughts and feelings are less likely to be expressed, such that specific screening for suicidal ideation has been suggested as a solution (Richa *et al.*, 2014).

Although there is a common misconception that talking about suicide may instigate or augment suicidal ideation, asking someone if they are having suicidal ideation, then listening without judgement, and without trying to 'fix' it, to their concerns, is the primary

way that non-professionals can help (O'Connor, 2021). And although suicide.net has been associated with suicides over many years, it is considered by its owners and many participants to be a highly valuable and beneficial community that largely helps prevent actual suicide by providing an easily accessible non-judgmental platform for discussion and support. For example, from suicide.net:

i have to give thanks to the creators of this site and the awesome people that have/had become a part of it. to the people that have passed may they rip, i strongly believe this site has saved many lives of people that might've been on the fence and would be devastated if this site were to ever go down

6.1.2 Suicide and the internet

The Samaritans, a UK charity aimed at providing emotional support for anyone in emotional distress, including at risk of suicide, state in their 2019 suicide statistics report that suicide-related internet use was found in 26% of deaths of under twenties and 13% of deaths of 20 to 24 year olds. A report on online self-harm and suicide content states "The internet can be an invaluable resource for individuals experiencing self-harm and suicidal feelings, however, it can also provide access to content that can be distressing and triggering" (Samaritans, 2020). A systematic review into the relationship between internet use, self-harm, and suicidal behaviour in young people similarly found that although there is a risk of harm from online behaviour in terms of normalisation, triggering, competition, and contagion relating to self-harm, there are also benefits in terms of crisis support, reduction of social isolation, delivery of therapy, and outreach, with young people in particular increasingly using social media to communicate their distress and seek support. It was recommended on this basis that clinicians working with young people with mental health issues should engage actively with discussions about related internet use (Marchant *et al.*, 2017).

An ethnographic investigation into the harms and benefits of online peer-support relating to self-harm of children and adolescents found that moves to eradicate online self-harm content may cause unintentional harm (Lavis and Winter, 2020). Benefits have been found in having access to others going through the same experience (Stommel, 2009), particularly for an issue that may attract social sanctions and negative judgement (White and Dorman, 2001), including overcoming barriers to access (Salem, Bogat and Reid, 1997). Health professionals must therefore take a more balanced view of self-harm and suicide websites, which "gave users access to important, socially valued identities, such as being understood, belonging to a community, and coping with their problems" (Baker and Fortune, 2008, p. 118).

In a comparison between telephone, synchronous internet chat, and asynchronous internet forum support it was found that the suicidal, who experience acute loneliness and unbearable mental pain, prefer the asynchronous support of an internet forum:

It seems that internet users who experience a critical mental state prefer to share their inner feelings with a group, in an asynchronous communication, rather than with an individual who may respond immediately. The group offers its participants an intense sense of affiliation and an experience of emotional support from other members (Gilat and Shahrar, 2007, p. 16)

A preference for chatting on the internet was also found in an investigation into the social dynamics of shame and youth suicide. New forms of communication such as internet fora, it was suggested, enable different kinds of expressive relations, and different social identities. In considering what can be done to address suicidal ideation, the young people interviewed in this research did not see traditional mental health services as a solution. Instead, they described the need for self-expression, and the feeling of being connected to others via a range of experiences that are not based on the typical youth-stressors of

success, goals, and achievement (Fullagar, 2003). Similarly, while research referred to above suggests on the basis of specific communication issues that those diagnosed as autistic should be medically screened for suicidal ideation, an investigation into the reasons behind their increased suicidal ideation finds it to be related to the loneliness that can result from being different in a neurotypical world – to not finding accessible ways to communicate and establish community (Hedley *et al.*, 2018). Bagatell argues that the internet-facilitated emergence of an autistic community has forced a less biomedical, more broad, consideration of "what it means to be social and how sociality is expressed" (Bagatell, 2010, p. 51). For example, from a participant diagnosed as autistic: "The computer is kind of like what sign language is for the Deaf. It's the autistic way of communicating" (Bagatell, 2010, p. 37).

6.1.3 Suicide and language style

In chapter one of the current study the many research findings about the relationship between language style and state of mind were summarised to support the current investigation of how metaphor use in an online community may be related to state of mind. In addition, studies considering language style specifically in conjunction with suicidal ideation (Wiltsey Stirman and Pennebaker, 2001; Pająk and Trzebiński, 2014; Egnoto and Griffin, 2016; O'Dea *et al.*, 2017) have found the following characteristics of suicidal discourse specifically (tabled below), and these are the primary focus in the current chapter.

Table 6-1 elements of language style found to be associated with suicidal ideation and suicide
lexical category **occurrence in suicidal ideation**

first-person singular pronouns	increased
first-person plural pronouns	decreased
pronouns referring to the other	decreased
range of pronouns	less diverse
anger words	increased
negative emotion words	increased

6.2 The suicide.net data

In this section the dimensions of suicide.net are described, followed by discussion of the keywords and key terms for suicide.net in comparison to a general web corpus, and in an intra-corpus comparison of discourse from participants who are more established in this community, against discourse from newer participants.

6.2.1 Collecting data from suicide.net

The Scrapy Python framework (Scrapy, 2021) was used to write web 'spiders' to collect data from the suicide.net forum. Every thread on the forum was crawled to pick up all posts on that thread, with each post saved as a separate record in a single JSON file, where JSON stands for JavaScript Object Notation, a common and useful data interchange format. The following data was collected for each post:

1. the id of the participant who made the post (anonymised)
2. the title of the thread the post was taken from
3. the id of the thread the post was taken from
4. the date the post was made
5. the text content of the post

6.2.2 Processing the data

The posts were sorted into ascending date order and converted to lower case. There are no healthcare professionals or professional administrators or moderators on suicide.net, and neither is there a composite anonymous participant. Therefore, all posts are retained in the data, with the participant id anonymised for privacy reasons.

The data was cleaned to remove formatting codes that are present in the text data. These are listed in the table below.

Table 6-2 formatting codes removed from cancer.net post text

code (removed)	represents
<code>\n</code>	new line
<code>\u00a0</code>	non-breaking space

The collected data was 'pickled' as a Python Pandas DataFrame, where a DataFrame is a powerful way of storing data that allows it to be queried as a relational database, and to be analysed using the many DataFrame data analysis methods. To ensure consistency, each subsequent analysis uses this same stored data.

6.2.3 suicide.net participants and posts: statistics and distribution over time

The suicide.net data of the current study consists of all forum posts from 18/03/2018, the first day of its existence at the independent web address sanctionedsuicide.com, up to 24/06/2020, the date the posts were collected. Earlier suicide.net data is no longer accessible on the web. The data is very diverse, with 424,468 posts made by 8,588 participants over a period of 28 months. The following table describes the dimensions of the data collected.

Table 6-3 suicide.net dimensions

count	description	referred to as
424,468	posts collected	
8,588	participants with at least 1 post	
1,025	participants with 1 post	
5,375	participants with <20 posts (33,298 posts)	suicide.net.LT20
3,213	participants with >19 posts (391,168 posts)	suicide.net.20
35	participants with > 1,000 posts	
49	average number of posts per active participant	

Individual suicide.net participants make around eight times as many posts on average in comparison to cancer.net participants, such that although suicide.net covers a time period of 28 months compared to 153 months for cancer.net, it contains around twice as many posts. This reflects the style of interaction on suicide.net, which is less formal, with many very short posts that form an ongoing conversation, and which may not make sense without their surrounding posts.

The distribution of participant posts on suicide.net meets the participation criterion of an online community, discussed in chapter one, in that it demonstrates "Frequent, regular, self-sustaining activity over time" (Herring, 2004, p. 361). In addition there are a number of extended threads on suicide.net which, in conjunction with distribution of posts by participant, meets the interaction criterion of an online community, which requires "Reciprocity, extended (in-depth) threads, core participants" (Herring, 2004, p. 361).

In the current study posts made by the more established participants who have made 20 or more posts are referred to as suicide.net.20, while posts made by newer participants who have made fewer than 20 posts are referred to as suicide.net.LT20.

6.3 Keywords and key terms

For the investigations in this section the Sketch Engine corpus analysis website was used to calculate keywords and key terms for suicide.net, where the keyness of terms, which is typically used to explore what a corpus is 'about', relates to their frequency in the corpus of focus compared to their frequency in a reference corpus. Inter-group comparisons between suicide.net and the general web corpus enTenTen18 provide insights into the particular concerns of suicide.net, while intra-group comparisons between more-established and less-established participants on suicide.net give an initial indication of the effect on participants of participation on the forum, in terms of both language style, which is considered in association with mental health, and discourse topics. In addition, intra-group comparison supports the use of data from the more established participants only in analysis in subsequent sections of community interpretative repertoires and their associated metaphors.

The tabled results, the key differences between the focus and reference corpora, are ranked in descending order of their score, which in this case is the Simple Maths Parameter: the frequency per million of a term in the focus corpus, divided by the frequency per million of a term in the reference corpus, with a value N , for which the default is 1, added to each frequency value before division, as a smoothing parameter. As the value of N is increased, so the focus is increasingly on less rare terms. As the value of N is decreased, so the focus is increasingly on more rare terms.

6.3.1 Keywords: comparison of suicide.net with enTenTen18

When suicide.net was compared with the general web corpus English Web 2018 (enTenTen18) the keywords, listed in the table below, were found to contain many acronyms that may be opaque to the out-group, fulfilling the structure criterion for

community, which refers to in-group jargon (Herring, 2004, p. 361). For example, when the score parameter $N=1$ (on the left of the table) the first keyword, with a very high score of 1,081, is *ctb*, an acronym for the metaphoric *catch the bus*, which in this community is a way of referring to suicide. *ctb* is sometimes described by external commentators as a euphemism for suicide, which is itself an example of the divide in understanding and perspective between the outgroup and suicide.net participants, who have no need for a euphemism for suicide: they use the term *suicide* approximately 58 times as often as it is used in the comparison general web corpus. On suicide.net *ctb* is also used in full as *catch the bus*, and the metaphor is often extended to include for example buying tickets for the bus, having a bus pass, waiting for the bus, and arrival and departure times for the bus.

it's a metaphor for the bus taking you to your final destination (death) (from suicide.net)

The second keyword is *sn*, an acronym for sodium nitrite, a valued suicide method. The acronyms *pph* and *ppeh* refer to the *peaceful pill handbook*, and *peaceful pill e-handbook*, written by a doctor and a lawyer (Nitschke and Stewart, 2021) to provide practical information about assisted suicide and euthanasia, which was first published in 2006.

When the score parameter $N=1,000$ so that the focus is on more common words (on the right of the following table), the highest scoring keyword is the first-person singular pronoun *i*, which is used approximately seven times as often on suicide.net as in the general web corpus, while the first-person singular pronouns *my*, and *me*, have the next highest score. In the following table metaphoric terms are underlined, and first-person singular pronouns are marked with an asterisk.

Table 6-4 Keywords, score parameter N=1 (left) and N=1,000 (right)

	keywords N=1	per M (focus)	per M (ref)	score		keywords N=1,000	per M (focus)	per M (ref)	score
1.	ctb	1393.4	0.3	1081.7		*i	40243.4	5649.9	6.2
2.	sn	1220.2	2.4	361.5		*my	9262.8	1579.3	4
3.	meto	211.9	0.1	202.6		*me	5882.6	864.1	3.7
4.	benzo	204.5	0.6	128.2		you	16656.1	5190.8	2.9
5.	antiemetic	135.2	0.2	110.2		do	11749.1	3496.8	2.8
6.	nembutal	104	0	101.8		feel	2892.4	399.1	2.8
7.	suicidal	418.9	3.7	88.8		just	4664.4	1048.4	2.8
8.	pph	98.3	0.2	80.5		not	15565.4	5191.4	2.7
9.	megathread	70.6	0	69.8		think	3526.1	720.3	2.6
10.	nitrite	139.6	1	69		want	3166.1	674	2.5
11.	metoclopramide	72.6	0.2	62.7		it	17740	6591	2.5
12.	ppeh	59.5	0	60.5		ctb	1393.4	0.3	2.4
13.	suicide	1339.7	23.5	54.7		would	5072.2	1579.8	2.4
14.	psych	130.7	1.4	53.9		so	5174	1697.2	2.3
15.	noose	85	0.6	52.3		suicide	1339.7	23.5	2.3
16.	hanging	160.2	2.3	49.5		if	5943.9	2049.1	2.3
17.	idk	81.5	0.8	46.3		life	2690.7	647.9	2.2
18.	tagamet	46.4	0.1	44.3		know	3487.6	1008.7	2.2
19.	painless	128	1.9	44		go	3948.5	1220	2.2
20.	anti-emetic	44.9	0.1	43.4		get	4349.9	1403	2.2

6.3.2 Key terms: comparison of suicide.net with enTenTen18

The key terms for suicide.net in comparison with the general web corpus, listed in the table below, provide further insight into the specific concerns of the forum. The first key term *survival instinct* represents the fear of being unable to complete an act of suicide because the more basic, embodied, *survival instinct* supersedes the self-conscious desire to die, potentially leaving the participant disabled, for example as represented by the key term *brain damage*, or subjected to mental health care against their will. It is partly with reference to the *survival instinct* that peaceful methods of suicide are sought that will not invoke it, for example "i chose a method so peaceful that the survival instinct will be almost non-existent", which is reflected in the metaphoric key term *peaceful journey*.

When the score parameter N=1,000 so that the focus is on more common terms (on the right of the table) the phrase *good luck* is the most increased in comparison to the reference corpus, occurring approximately 20 times as often. As well as communicating support and good wishes, this reflects the sense on suicide.net that since suicide is not a part of hegemonic culture, which is regulated in many aspects, participants are engaging with the unknown in terms of the potential act of suicide and potential outcomes. The metaphoric key term *other side* extends the metaphorical conceptualisation of life, and suicide, as a journey, to include what happens after suicide, for example: "i have this image of a door that i would just walk through if i could, and on the other side is not-life". Talk of *mental health* and *mental illness* are also dominant.

Table 6-5 key terms, score parameter N=1 (left) and N=1,000 (right)

key terms N=1	per M (focus)	per M (ref)	score	key terms N=1,000	per M (focus)	per M (ref)	score
1. survival instinct	100.7	0.1	91.9	good luck	182.9	9.4	1.2
2. sodium nitrite	89.0	0.1	82.2	mental health	164.1	22.0	1.1
3. exit bag	73.6	0	74.6	survival instinct	100.7	0.1	1.1
4. full suspension	76.9	0.1	71.1	mental illness	98.6	5.8	1.1
5. partial suspension	63.7	0	64.1	sodium nitrite	89.0	0.1	1.1
6. partial hanging	44.0	0	44.9	long time	119.0	29.8	1.1
7. sn method	40.4	0	41.4	only thing	90.9	11.3	1.1
8. peaceful pill	39.1	0	40.1	full suspension	76.9	0.1	1.1
9. stat dose	35.6	0	36.6	exit bag	73.6	0	1.1
10. inert gas	44.4	0.3	34.6	brain damage	65.3	1.3	1.1
11. lethal dose	34.1	0.2	29.4	partial suspension	63.7	0	1.1
12. brain damage	65.3	1.3	28.4	good idea	61.8	12.9	1
13. night night	28.3	0	28.3	same thing	63.0	15.4	1
14. suicide attempt	34.2	0.4	25.2	last time	58.0	11.7	1

15. suicide note	26.8	0.2	23.8	only way	59.6	13.3	1
16. much pain	36.6	0.7	22.3	inert gas	44.4	0.3	1
17. <u>peaceful</u> <u>journey</u>	21.4	0	22.2	partial hanging	44.0	0	1
18. self harm	23.7	0.1	21.6	own life	46.6	3.8	1
19. suspension hanging	20.5	0	21.5	carbon monoxide	43.8	2.1	1
20. peaceful death	21.3	0	21.5	other side	58.0	15.8	1

6.3.3 Comparison of different suicide.net groups with enTenTen18

To consider the differences in language between more established participants on suicide.net who have made 20 or more posts (suicide.net.20, on the left of the table below) in comparison with less established participants who have made fewer than 20 posts (suicide.net.LT20, on the right of the table), in this analysis keywords for each of these suicide.net groups were calculated with the general web corpus as the reference corpus. In order to focus on more common words the score parameter N was set to 1,000.

It can be seen from the results table below that although the keywords and the language style they represent are very similar for both suicide.net groups, the language style of the more established participant group suggests slightly better mental health. For example, the first-person singular pronoun *i* has a score of 6, in comparison to 7.7 for the less established group, and the first-person singular pronouns *my*, and *me*, are similarly less prevalent in the more established group. In addition, suicide.net.20 has the other pronoun *you* as a keyword, while suicide.net.LT20 has no pronouns referring to the other but does have the additional first-person singular pronoun *myself*.

Table 6-6 keywords, more established participants (left), and newer participants (right)

N= 1,000	suicide .net.20	per M (focus)	per M (ref)	score	suicide.net .LT20	per M (focus)	per M (ref)	score
1.	*i	39167.60	5649.90	6.000	*i	50226.10	5649.90	7.700
2.	*my	8868.90	1579.30	3.800	*my	12916.80	1579.30	5.400
3.	*me	5696.00	864.10	3.600	*me	7613.60	864.10	4.600
4.	you	17163.90	5190.80	2.900	feel	3686.60	399.10	3.300
5.	do	11709.50	3496.80	2.800	just	5181.00	1048.40	3.000
6.	just	4608.80	1048.40	2.700	do	12117.80	3496.80	2.900
7.	feel	2806.80	399.10	2.700	want	3783.80	674.00	2.900
8.	not	15610.90	5191.40	2.700	think	3680.50	720.30	2.700
9.	think	3509.50	720.30	2.600	not	15144.50	5191.40	2.600
10.	it	17852.30	6591.00	2.500	*myself	1827.60	86.00	2.600
11.	want	3099.60	674.00	2.400	so	5645.80	1697.20	2.500
12.	ctb	1407.80	0.30	2.400	would	5214.70	1579.80	2.400
13.	would	5056.90	1579.80	2.300	but	8132.50	2822.40	2.400
14.	if	6009.10	2049.10	2.300	try	2413.00	434.80	2.400
15.	suicide	1347.40	23.50	2.300	life	2880.40	647.90	2.400
16.	so	5123.10	1697.20	2.300	get	4640.30	1403.00	2.300
17.	sn	1244.10	2.40	2.200	know	3714.60	1008.70	2.300
18.	life	2670.30	647.90	2.200	go	4177.40	1220.00	2.300
19.	know	3463.20	1008.70	2.200	it	16697.50	6591.00	2.300
20.	go	3923.70	1220.00	2.200	like	4103.30	1196.30	2.300
21.	get	4318.50	1403.00	2.200	ctb	1260.40	0.30	2.300
22.	like	3803.20	1196.30	2.200	really	2151.70	403.00	2.200

6.3.4 Comparison of keywords between suicide.net participant groups

To further consider the effect on mental health of suicide.net as a community, for the investigation in this section a comparison was made between suicide.net.20 (more established participants) and suicide.net.LT20 (less established participants), each with the other as the reference corpus. The following table lists the keywords for suicide.net.20 with suicide.net.LT20 as the reference corpus on the left, and the keywords for suicide.net.LT20 with suicide.net.20 as the reference corpus on the right. The score parameter N was set to 1,000 for this analysis so that the focus is on more common words.

It can be seen from the results table below that suicide.net.20 has more other, and group, references, with *you*, *your*, *we*, *they*, and *their* as the top keywords, with *us*, *yourself*, *u*, and *his* also keywords (marked in bold), and no first-person singular pronouns are keywords for this group. Suicide.net.LT20 in comparison has the first-person singular pronouns *my*, *myself*, *me*, and *i* as its top keywords (marked with an asterisk), and no references to the other. This shows that more established suicide.net participants have a language style more associated with better mental health in comparison to less established participants.

Table 6-7 keywords for suicide.net.20 (left) and suicide.net.LT20 (right) with each other as reference corpus

N=	suicide	per M	per M	score	suicide	per M	per M	score
1,000	.net.20	(focus)	(ref)		.net.LT20	(focus)	(ref)	
1.	you	17163.90	11937.80	1.400	*my	12916.80	8868.9	1.400
2.	your	4265.80	2936.70	1.300	*myself	1827.60	1144.90	1.300
3.	we	2206.20	1629.40	1.200	*me	7613.6	5696.000	1.300
4.	they	4303.20	3377.60	1.200	*i	50226.10	39167.60	1.300
5.	their	1329.30	923.10	1.200	feel	3686.600	2806.80	1.200
6.	people	3448.60	2809.30	1.200	anyone	1100.80	769.00	1.200
7.	us	832.60	569.90	1.200	year	1623.80	1227.70	1.200
8.	say	2823.70	2286.30	1.200	want	3783.80	3099.60	1.200
9.	may	852.30	597.50	1.200	thank	757.10	514.40	1.200
10.	yes	661.90	439.00	1.200	hi	295.30	117.40	1.200
11.	n	718.40	501.00	1.100	thanks	647.80	427.00	1.200
12.	sn	1244.100	999.80	1.100	friend	995.70	743.80	1.100
13.	thread	590.80	420.10	1.100	family	898.50	659.70	1.100
14.	yourself	576.40	409.20	1.100	try	2413.000	1990.80	1.100
15.	u	342.00	203.80	1.100	*im	602.80	409.10	1.100
16.	if	6009.10	5336.10	1.100	plan	855.40	636.30	1.100
17.	case	512.10	367.20	1.100	rope	423.50	266.40	1.100
18.	death	1036.00	841.00	1.100	since	856.90	654.20	1.100
19.	his	667.90	511.60	1.100	any	1843.90	1541.70	1.100
20.	lol	306.50	184.20	1.100	really	2151.70	1818.20	1.100

6.3.5 Comparison of key terms between suicide.net participant groups

In order to further explore the differences in language between more established and newer participants on suicide.net, the same comparisons carried out for keywords in the previous section were carried out for key terms. The top key term for more established participants, on the left in the table below, is *good luck*, which as discussed previously is typically addressed towards another participant. Key terms for this group also include the metaphoric phrases *peaceful journey* and *safe journey*, as well as references to community identity and topics, such as *goodbye thread*, *peaceful pill handbook*, and *pro choice*.

The more recent participants in comparison, on the right in the table below, use introductory phrases such as *first time*, *first post*, *hello everyone*, and *new member*, demonstrating that new participants tend to formally introduce themselves on the forum, reinforcing interpretation of it as a community such that social norms must be met. Typical terms for newer participants also include the specific personal health references *self harm*, and *mental health*. There are no metaphoric key terms for newer participants.

Table 6-8 key terms for suicide.net.20(left) and suicide.net.LT20 (right) with each other as a reference

N=	suicide.net.	per M	per M	score	suicide.net.	per M	per M	score
1,000	20	(focus)	(ref)		LT20	(focus)	(ref)	
1.	good luck	188.50	130.50	1.100	first time	128.60	70.00	1.100
2.	stat dose	38.10	12.50	1.000	partial suspension	110.00	58.70	1.000
3.	sodium nitrite	90.80	73.00	1.000	only thing	135.80	86.00	1.000
4.	peaceful pill	40.60	25.70	1.000	partial hanging	80.20	40.10	1.000
5.	pill handbook	27.00	13.60	1.000	first post	48.80	13.10	1.000
6.	goodbye thread	17.90	4.90	1.000	last year	80.90	52.50	1.000
7.	peaceful pill handbook	24.90	12.50	1.000	hello everyone	31.40	4.50	1.000
8.	survival instinct	102.00	88.90	1.000	carotid artery	40.50	17.00	1.000
9.	old age	20.40	9.50	1.000	last time	76.80	56.00	1.000

10. <u>peaceful journey</u>	22.40	12.10	1.000	exit bag	92.60	71.60	1.000
11. method introduction	11.50	1.50	1.000	self harm	39.70	22.00	1.000
12. full understanding	11.90	1.90	1.000	high school	51.10	34.20	1.000
13. suicide sodium nitrite	11.50	1.50	1.000	slip knot	26.80	10.80	1.000
14. peaceful method	15.60	5.70	1.000	long time	133.50	117.50	1.000
15. law enforcement	15.50	5.70	1.000	carbon monoxide	57.10	42.40	1.000
16. good idea	62.80	52.60	1.000	sn method	53.30	39.00	1.000
17. good thing	36.00	26.10	1.000	mental health	178.50	162.50	1.000
18. <u>safe journey</u>	16.80	7.20	1.000	long story	25.30	11.70	1.000
19. inert gas	45.30	35.90	1.000	new member	18.90	5.50	1.000
20. pro choice	17.20	8.70	1.000	night method	28.70	15.30	1.000

6.3.6 Section summary: Keywords and key terms

It was shown that the keywords and key terms of suicide.net demonstrate a very high focus on the self in comparison to a general web corpus, and on specific topics relating to suicide, including the metaphoric *ctb*. In addition, comparison of keyness between suicide.net.20 and suicide.net.LT20 found that newer participants use more self references in comparison to more established participants, with the latter using more other, and group, references. The language of more established suicide.net participants is of a style that has been associated with better mental health in comparison to more recent suicide.net participants, but not in comparison to the general web corpus. When the focus was on more rare key terms, insights were gained into the particular language and concerns of this community.

The investigations in this section have shown that the suicide.net.20 data is less focused on the current lived experience of a particular participant, and thus more reflective of community interpretative repertoires. For this reason, where typical community discourse

is sought in subsequent investigations in the current chapter, the suicide.net.20 data is used.

6.4 Interpretative repertoires

In this section Latent Dirichlet Allocation (LDA) topic modelling (section 0) is used with a range of settings, to locate the interpretative repertoires (section 1.5.2) and ideological dilemmas (section 1.5.3) of suicide.net. The focus is on potentially metaphoric terms, with further potentially metaphoric terms in suicide.net extracted via the word vector similarity method used with cancer.net (section 2.9), which locates word in the data that are used similarly in general web discourse, including metaphoric use. The suicide.net.20 data is used in this section.

6.4.1 Using LDA topic modelling to find interpretative repertoires

6.4.1.1 LDA with noun lemmas, trigrams, max_df=0.1, min_df=2

In this LDA analysis only noun lemmas were included, and from those bigrams and trigrams were formed. In order to focus on those more rare terms more likely to be metaphoric (Philip, 2010), the text data was vectorised with a max_df setting of 0.1, so that terms were only included that are present in 10% or fewer of posts. In order to exclude idiosyncratic terms, a min_df setting of 2 was used, so that terms were only included that are present in at least two posts.

The 25 topics in the LDA output each have a clear focus, which is reflected in the (high) 0.69 coherence score, where the coherence score indicates how likely it is that the terms in a topic do occur together in any particular document. The coherence of topics is another indication that suicide.net.20 is a community in that it has dominant and distinct common concerns.

While the first 11 LDA topics were found to be more focused, and less general, each of the 25 topics provides different insights into the concerns and language of suicide.net participants, and potentially of people who feel suicidal in general. The metonymic term *lifer* in topic one, which appears to be about the controversy around suicide as a valid choice, represents the outgroup of those who deny the validity of suicide as a choice, and who seek to shut down suicide.net; *lifer* also has connotations of the dogma of extreme right-wing and religious affiliation, including anti-abortion activism. Topic two contains the potentially metaphoric noun *war*, although in relation to the existential considerations of topic two *war* may be used literally in this context, while *fight* in topic three is perhaps more likely to be metaphoric since topic three appears to relate to difficult home life. Topic 25 contains the metaphoric *ctb* and the potentially metaphoric *journey*, while topic 24 contains an expansion of metaphoric *ctb*, in the term *catch_bus*.

Table 6-9 LDA topics, noun lemmas, trigrams, with max_df=0.1

topic	topic is about?	terms listed in order of prevalence in topic
1.	controversy around suicide as a valid choice	person, society, choice, accept, agree, fact, control, reason, view, respect, action, opinion, choose, point, force, response, matter, act, perspective, individual, argument, majority, seek, support, effort, truth, decision, blame, circumstances, <u>lifer</u>
2.	existential considerations	world, live, hell, exist, god, human, suffering, bear, create, existence, animal, nature, soul, earth, afterlife, universe, reality, religion, suffer, theory, control, make_sense, meaning, consciousness, part, matter, belief, energy, planet, power, torture, nothingness, faith, <u>war</u> , reincarnation, bible
3.	difficult home life	year, live, parent, child, kid, age, mother, mom, family, leave, stay, abuse, dad, grow, move, mum, home, wife, brother, school, home, father, adult, blame, childhood, sister, bring, <u>fight</u> , lose, husband, cry, cat, daughter, son, hate, baby
4.	legal considerations around suicide	suicide, case, state, lead, report, commit_suicide, follow, term, involve, act, mention, reason, prevent, result, fact, include, regard, base, article, assume, factor, refer, knowledge, study, avoid, law, form, point, evidence, victim
5.	medical treatment and conditions	doctor, hospital, illness, health, med, treat, therapy, treatment, therapist, euthanasia, psychiatrist, patient, force, psych_ward, section, year, call, disorder, prison, diagnose, cure, condition, give, medication, diagnosis, depression, nurse, bpd, visit, surgery
6.	drugs and suicide	drug, effect, dose, benzo, pill, overdose, alcohol, mg, side_effect, symptom, dosage, medication, heroin, prescribe,

		combine, increase, tolerance, coma, inject, combination, knock, seizure, od, opiate, opioid, sedative, cocktail, antidepressant, morphine, injection
7.	issues that prompt suicidal thoughts	work, money, job, pay, live, problem, afford, start, college, spend, career, degree, quit, goal, debt, lose, disability, fix, motivation, government, opportunity, class, year, skill, benefit, interest, school, give, living
8.	practical concerns around suicide	minute, bag, test, gas, air, nitrogen, breathe, oxygen, mask, exit_bag, breath, seal, line, tank, remove, design, helium, sise, set, acid, fill, open, regulator, carbon_monoxide, co2, put, measure, level, flow, breathing
9.	arrangements for suicide in terms of the place it is done	place, find, put, leave, car, room, run, drive, house, walk, home, burn, hotel, light, sit, dog, move, area, charcoal, city, travel, stay, throw, smoke, tent, apartment, wood, door, window, tree
10.	rationale for experience of suicidal ideation	experience, thought, feeling, mind, depression, anxiety, dream, moment, deal, thing, bring, relate, remember, part, emotion, memory, sense, struggle, cope, escape, trauma, realise, stress, desire, trap, describe, share, overcome, relief, trigger
11.	writing suicide notes	write, send, note, find, give, police, contact, phone, call, email, check, number, account, message, leave, explain, sign, letter, detail, show, service, trust, record, add, paper, photo, address, text, depend, internet
12.	the changing nature of life	thing, happen, change, enjoy, end, lot, play, learn, regret, sort, forget, matter, part, movie, moment, remember, turn, suppose, past, watch, game, put mistake, music, song, mess, improve, bit, point, make_mistake
13.	dates and times	day, week, wait, sleep, month, today, start, wake, night, hour, bed, stop eat, tomorrow, morning, yesterday, trip, birthday, spend, cry, stay, tonight, couple, return, weekend, nightmare, plan, rest, remember, shower
14.	problems of suicide related to those who are bereaved by it	love, friend, care, family, hurt, leave, lose, meet, hate, miss, deserve, relationship, guilt, loss, support, fault, promise, burden, heart, stranger, grief, christmas, give, pet, world, isolate, break, abandon, understand, happiness
15.	issues around talking about suicide	talk, understand, person, speak, situation, word, hear, listen, give, place, call, reason, stop, group, judge, advice, agree, explain, support, conversation, vent, opinion, problem, interact, voice, point, story, struggle, attention, show
16.	negative thoughts about death and suicide	die, death, pain, kill, body, fear, suffer, end, brain ,suffering, cancer, heart, endure, stop, accident, survival_instinct, agony, coward, organ, heart_attack, pray, torture, save, condition, rest, imagine, disease, documentary, scare, misery
17.	sn as a suicide method	sn, drink, meto, taste, vomit, antemitic, bottle, stomach, water, mix, amount, hour, throw, poison, gram, recommend, salt, swallow, food, minute, vomiting, liquid, consume, cyanide, puke, fast, min, regimen, solution, ingest
18.	hanging as a suicide method	hang, pass, rope, cut, suspension, hanging, neck, hand, drop, practice, position, pull, weight, head, pressure, stand, tie,

		noose, hold, belt, knot, blood, back, finger, push, arm, black, press, break, choke
19.		fuck, shit, man, guy, woman, face, partner, hate, girl, give, lie, suck, turn, happen, eye, story, start, show, sex, joke, laugh, date, watch, lot, fun, touch, bullshit, ruin, ass, point
20.	non-chemical methods	jump, fall, survive, gun, head, hit, train, shoot, chance, damage, foot, bridge, shot, point, drown, shotgun, firearm, land, story, side, aim, impact, height, picture, video, fire, injury, load, blow, area
21.	how to use the forum	post, read, thread, forum, site, member, pm, information, link, op, comment, user, reply, info, discuss, ss, search, page, question, share, join, suggest, chat, rule, discussion, ask_question, guide, megathread, ban, topic
22.	methods and potential outcomes	method, attempt, find, fail, option, research, choose, risk, failure, save, succeed, choice, chance, courage, fail_attempt, success, brain_damage, guarantee, rescue, plan, plenty, prefer, prepare, idea, trouble, rush, planning, discover, suggestion, avoid
23.	buying chemicals for suicide	buy, order, country, source, sell, list, receive, uk, exit, book, obtain, seller, product, update, check, purchase, website, package, ship, access, custom, mention, confirm, scam, pph, deliver, arrive, market, price, usa
24.	talking around reasons for suicide	sound, guess, lot, idea, gon, nt, stuff, hear, worry, answer, bit, type, kind, wanna, figure, bother, issue, question, stick, lol, problem, suppose, find, imagine, expect, happen, crap, <u>catch_bus</u>
25.	ctb	<u>ctb</u> , hope, plan, decide, peace, luck, decision, reason, end, hope_find_peace, find, hug, comfort, find_peace, <u>journey</u> , mind, miss, wish_peace, change_mind, future, choose, hold, situation, wish_luck, deserve, side, travel, si, option, goodbye

6.4.1.2 Comparison of potentially metaphoric topics from different LDA settings

In the topic modelling analysis in the previous section noun lemmas were included that are present in 10% or fewer of suicide.net.20 posts. In the current section, to further explore the metaphoric interpretative repertoires that were found, results are compared when adjectives and verbs are also included in the analysis, and for different max_df settings representing terms that are present in different proportions of suicide.net posts.

It can be seen from the results summary table below that *ctb* and *journey* were found to be typically used together when the max_df setting was greater than 0.04, but at max_df=0.04 (the term is present in 4% or fewer of posts) they split into two separate

potentially metaphoric topics, giving the insight that they do different work, and also, from the repertoire of terms they are used with, what that work is.

The presence of *bus* alongside *ctb* when max_df was set to 0.05 or less supports the interpretation of *ctb* as an active metaphoric framing, as does the metaphoric transport term *boat* in the same topic, which is typically used in the metaphoric phrase *the same boat* to communicate solidarity and understanding. The interpretative repertoire associated with *ctb* appears to be related to planning, including references to dates and occasions. The related interpretative repertoire associated with *journey* shows that it is typically used to wish another participant (*yr=your*) *peace* and *good_luck* on their *journey*, i.e. when they are about to *catch the bus*. When adjectives were included in the analysis, use of the metaphoric term *other side* in conjunction with *ctb* and *journey* became apparent.

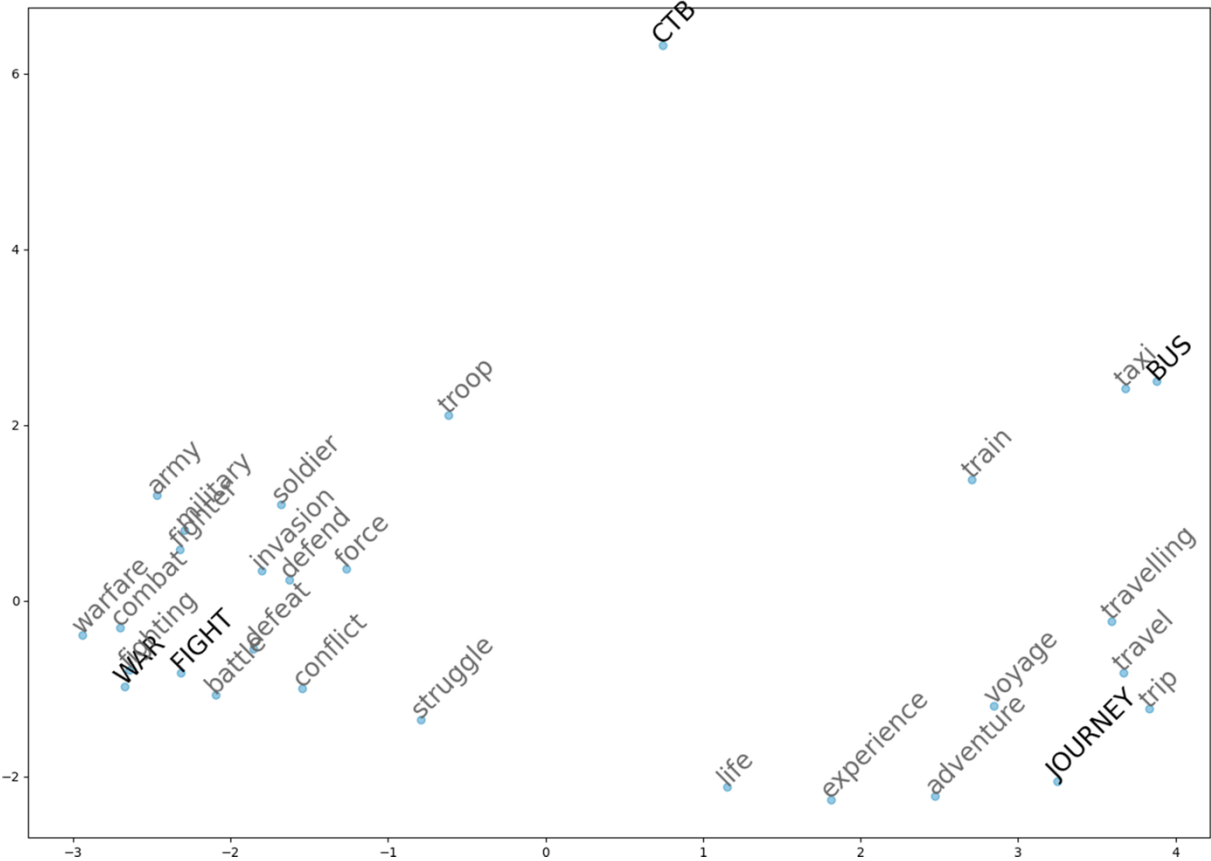
Table 6-10 potentially metaphoric topics from different LDA settings

max_df	lemmas	potentially metaphoric topics	coherence score
0.1	NOUN	<u>ctb</u> , hope, plan, decide, peace, luck, decision, reason, end, hope_find_peace, find, hug, comfort, find_peace, <u>journey</u> , mind, miss, wish_peace, change_mind, future, choose, hold, situation, wish_luck, deserve, side, travel, si, option, goodbye	0.69
0.1	NOUN, VERB	<u>ctb</u> , plan, leave, reason, wait, decide, family, date, thought, cat, end, prepare, worry, birthday, guess, funeral, holiday, christmas, impulse, delay, rush, stick, urge, chance, set, hold, month, leave_note, say_goodbye, future	0.70
0.1	NOUN, VERB, ADJ	<u>ctb</u> , hope, find, peace, decide, sound, good_luck, hug, glad, plan, hope_find_peace, decision, comfort, miss, find_peace, partner, sorry_hear, thought, ready, future, change_mind, <u>journey</u> , rush, luck, sad, give, wish_peace, peaceful, bring, <u>other_side</u>	0.65
0.05	NOUN	peace, <u>ctb</u> , attempt, place, luck, plan, note, chance, hug, idea, thought, failure, date, si, <u>journey</u> , bit, <u>bus</u> , comfort, goodbye, courage, success, mind, worry, travel, impulse, strength, planning, ppl, birthday, buddy	0.76
0.04	NOUN	<u>ctb</u> , week, month, plan, today, end, bit, date, idea, <u>bus</u> , tomorrow, couple, yesterday, tonight, birthday, chance, worry, place, boat, holiday, weekend, christmas, update, ppl, trouble, excuse, summer, step, march, january	0.76
0.04	NOUN	peace, luck, amp, bit, place, hug, <u>journey</u> , rest, lol, kind, side, comfort, goodbye, strength, light, travel, tho, stuff, dmso, plenty, plant, yr, step, load, relief, seed, transition, kindness	0.76

6.4.2 Using word vectors to explore metaphoric language

In this section the potentially metaphoric terms discovered with LDA topic modelling were used to find other potentially metaphoric nouns in suicide.net, based on the similarity of the word vectors that represent their typical use in general web discourse. Terms of focus are capitalised in the following plot to make them easier to find. This analysis is based on lemmas of the relevant nouns.

Figure 6-1 plot of potentially metaphoric terms (capitalised) and other similar terms (similarity > 0.6) in suicide.net.20



The nouns most similar to the potentially metaphoric terms *ctb*, *bus*, *journey*, *fight*, and *war*, in suicide.net (similarity > 0.6) are listed in the following table in descending order of similarity.

Table 6-11 suicide.net.20 terms with word vector similarity > 0.6 with potentially metaphoric terms found via LDA topic modelling

potentially metaphoric term	similar noun in suicide.net	similarity (>0.60)
bus	taxi	0.72524
bus	train	0.64788
journey	trip	0.64983
journey	voyage	0.64639
journey	adventure	0.63869
journey	life	0.63291
journey	experience	0.60103
fight	battle	0.76878
fight	fighter	0.65026

fight	defeat	0.61585
fight	combat	0.60842
fight	struggle	0.60828
war	military	0.69272
war	warfare	0.67774
war	conflict	0.67472
war	army	0.67405
war	troop	0.65831
war	soldier	0.65274
war	battle	0.65086
war	force	0.61249
war	invasion	0.60256

The actual metaphoricity of these terms in suicide.net.20 is considered in the next section.

6.4.3 Section summary: interpretative repertoires

Distinct interpretative repertoires were found in suicide.net.20 which provide a wide-ranging insight into the concerns of this community, while the high coherence value of the topics found shows that the community has clear common concerns. The potentially metaphoric terms *ctb* and *journey* were found in the same interpretative repertoire at some term max_df (density) settings, but not others, suggesting that they do different but related work within suicide.net, such that their presence together may represent an ideological dilemma, perhaps the tension between actual suicide, and membership of a suicide community. Potentially metaphoric terms found in the interpretative repertoires were further explored to find other nouns in the data with which they have a high word vector similarity, nouns which may be insufficiently prevalent, or too prevalent, to be found according to the max_df settings used in the LDA topic modelling investigations. In the next section all of these potentially metaphoric terms are evaluated for actual metaphoricity in this context.

6.5 Evaluating potential metaphor

As discussed in chapter three, MIPVU (Steen *et al.*, 2010) is used in the current study to determine the actual metaphoricity of potentially metaphoric terms. As with the same investigation for cancer.net, the Macmillan online dictionary is used to determine the contemporary meaning of each term as a noun. While MIPVU is a method for identifying potential metaphor, it is used here as a formal method to consider the actual metaphoricity of terms, with 20 random instances of the lemma of each term as a noun sampled as a Sketch Engine concordance analysis to determine the proportion in which it is used metaphorically in the suicide.net context. The phrase *catch the bus* is considered as a simple phrase, so that similar terms such as *catching the bus* are also identified. The following table summarises into themes the potentially metaphoric terms that are evaluated in the current section.

Table 6-12 potentially metaphoric terms and their similarly used nouns from suicide.net.20

term	similarly used nouns from suicide.net.20
ctb	
catch the bus	
bus	taxi, train
journey	experience, life, adventure, voyage, trip
other side	
fight	struggle, combat, defeat, fighter, battle
war	invasion, force, battle, soldier, troop, army, conflict, warfare, military

In the results tables of the following sections the following attributes are summarised for each potentially metaphoric term:

1. the number of instances of the term
2. the density of the term per one million tokens
3. the proportion of analysed occurrences that are metaphoric in their context

Metaphoric examples are also provided to build a sense of the work that these metaphors do in this context, while examples of non-metaphoric use are crossed out. In addition, for the primary potentially metaphoric terms being explored, i.e. those that were identified by their presence in interpretative repertoires, their density in suicide.net.20 is compared to their density in suicide.net.LT20. All other terms are investigated solely with regard to suicide.net.20.

6.5.1 journey and related nouns

The metaphoric term *other side* used in the same interpretative repertoire as *journey* was also included in this analysis. It can be seen from the following results table that *journey* was found to always be metaphoric in this context and to be used approximately 1.5 times as often in suicide.net.20 as it is in suicide.net.LT20, suggesting that it is a community metaphor, i.e. a metaphor that is characteristic of the community and has been adapted to do particular work within it. None of the potentially metaphoric nouns related to *journey* were found to be used predominantly metaphorically in this context, while *other side* is predominantly used metaphorically for less established but not for more established participants.

Table 6-13 journey and related terms evaluation of metaphoricity

term	count	density	metaphoric instances	metaphoric examples
journey suicide.net.20	2,676	109.15	20/20	emotions and other synapses tagging along for the cosmic journey
journey suicide.net.LT20	190	71.85	20/20	it's a catch 22 i wish you a pleasant journey whether you stay or go
experience	9,990	407.49	0/20	came out even more scarred from the hospital experience
life	65,463	2670.25	0/20	i was kicked out of their house which threw my life off the otherwise stable rails
adventure	254	10.36	0/20	death is the only adventure that i can get excited about

voyage	68	2.77	3/20 (15%)	if i could get my hands on a bottle of n, i would make my voyage tonight a voyage to antarctica bon voyage
trip	1,514	61.76	9/20 (45%)	i have also heard death is the best trip, that's why they save it for last i plan to go on a trip from morocco to mexico to buy it
other side suicide.net.20	1,444	58.9	10/20 (50%)	you have to be ready to face the unknown that lies on the other side i anchored it to a door knob and threw the noose over to the other side of the door
other side suicide.net.LT20	138	52.18	12/20 (60%)	we're institutionalized to life. we don't know, we don't want to know, and we will not believe what's on the other side and then to the other side starting from one end of the cut

6.5.2 ctb and related nouns

The following results table shows that *ctb* is always metaphoric in this context. Use of the related term *bus* is sometimes literal, but always metaphoric when used in conjunction with *catch*. The terms *taxi* and *train* in comparison are never used metaphorically in this context. In addition, *ctb* is used in a higher proportion of posts in suicide.net.20 than it is in suicide.net.LT20.

Table 6-14 ctb and related terms evaluation of metaphoricity

term	count	density	metaphoric instances	metaphoric examples
ctb suicide.net.20	34,437	1,404.69	20/20	have wanted to ctb since i was six
ctb suicide.net.LT20	3,326	1,257.72	20/20	we all would rather they felt better than ctb
bus suicide.net.20	2,330	9.04	14/20 (70%)	i hope she went peacefully and hopefully we can all catch that same bus down the line taking a bus or something out to the middle of the desert
bus suicide.net.LT20	204	77.14	14/20 (70%)	i think i will give it another go today, i have my bus pass:) being ridiculed and laughed at on the bus

catch the bus suicide.net.20	826	33.69	20/20	i'm going to be seen as the villain no matter what for catching the bus
catch the bus suicide.net.LT20	101	38.19	20/20	i'm relatively comfortable with the idea of not catching the bus
taxi	157	6.4	0/20	he insisted on paying for a taxi
train	2,315	135.22	0/20	observe how fast the train moves past the spot

6.5.3 fight and related nouns

The following results table shows that the nouns *fight*, *struggle*, *fighter*, and *battle*, are used predominantly metaphorically in this context, while the other similar nouns are not. However, there is also significant literal use of *fight*, for example as a method of suicide: "i'd just cross into palestine and start a fight. you may die horribly, but death is death if you're gonna be arrested for suicide". It can also be seen that density of *fight* is similar in suicide.net.20 and suicide.net.LT20: its use does not increase with increased participation on the forum.

Table 6-15 fight and related terms evaluation of metaphoricity

term	count	density	metaphoric instances	metaphoric examples
fight suicide.net.20	899	36.67	15/20 (75%)	the incredibly entertaining fight between good and evil on earth you're better off getting into a fight with a gang or something
fight suicide.net.LT20	104	39.33	15/20 (75%)	we had a big emotional talk/fight i did some amateur fights in the past
struggle	1,907	77.75	20/20	facing death is a big struggle, despite all my desire to ctb
combat	61	2.49	0/20	the winner of the physical combat
defeat	69	2.81	0/20	each new day feels like another defeat
fighter	67	2.73	14/20 (70%)	this isn't a cancer fighter lose your battle bullshit how can he be a male fighter pilot and some girl the next moment?
battle	546	22.27	18/20 (90%)	i believed that i was in a mental battle against the queen

~~the same as people dying in battle~~

6.5.4 war and related terms

The following results table shows that although *war* is used metaphorically in the suicide.net context, it is predominantly used literally. Of the other potentially metaphoric terms related to *war*, only *invasion*, and *force* are used predominantly metaphorically in this context.

Table 6-16 war and related terms analysis of metaphoricity

term	count	density	metaphoric instances	metaphoric examples
war suicide.net.20	804	32.8	6/20	i don't want to have a war with you world war 1
war suicide.net.LT20	78	29.5	6/20	i made a war zone in my apartment will you blame a person for not fighting in war to achieve something?
invasion	23	0.94	15/20 (75%)	the rummaging, ransacking, and invasion of privacy that suicide can lead to i have ptsd from being kidnapped in a home invasion
force	1,129	46.05	12/20 (60%)	this is the driving force behind alcoholism and addiction the blade might not come down with enough force
soldier	169	6.89	0/20	an american revolutionary war soldier
troop	16	0.65	0/6	feed back enemy troop positions to hq
army	181	7.38	0/20	avoid capture by an army
conflict	290	11.83	10/20 (50%)	values that are in conflict insane debts that were put on germany after the conflict
warfare	15	0.61	0/20	history of warfare
military	156	6.36	0/20	joining the military

6.5.5 Comparative densities of metaphor for different participant groups

In the previous sections, for each of the primary metaphoric terms identified in interpretative repertoires in section 6.4 above, the density of that term as a noun lemma was compared between more established and less established suicide.net participants.

This comparative information is summarised in the following table, and in addition the comparative densities for the nouns *battle*, and *struggle* are included, since they were found to be metaphoric in 90% or more of occurrences, and to have a relatively high density. Although *struggle* was present in two interpretative repertoires, it was not initially identified as metaphoric in that context; the nouns *struggle* and *battle* have both been identified as metaphoric through the word vector similarity method combined with the metaphor evaluation of the current section.

Table 6-17 density of metaphoric terms for different segments of the cancer.net community

metaphoric term	% metaphoric	suicide.net.20 density	suicide.net.LT20 density	suicide.net.20 / suicide.net.LT20
journey	100	109.15	71.85	1.52
ctb	100	1,404.69	1,257.72	1.12
battle	90	22.27	22.69	0.98
fight	75	36.67	39.33	0.93
catch the bus	100	33.69	38.19	0.88
struggle	100	77.75	93.02	0.84

The summary table above, which is ordered by change in density between suicide.net.LT20 and suicide.net.20, shows that the *journey* metaphor is more prevalent in the suicide.net.20 data of more established participants; *journey* also has the highest change in density between the two participant conditions. The metaphoric acronym *ctb*, which is the most used metaphor for both sets of data, is also more prevalent in suicide.net.20 than in suicide.net.LT20. In comparison, the expanded metaphoric phrase *catch the bus* is more prevalent in suicide.net.LT20 than in suicide.net.20, as are the metaphoric terms *battle*, *fight*, and *struggle*.

In considering how to group these metaphoric terms into themes reference is made to the interpretative repertoires described in section 6.4. The nouns *journey*, and *fight*, each appear in one interpretative repertoire, while *struggle* appears in two interpretative repertoires, with none of these three nouns coinciding in the same interpretative

repertoire. The metaphoric phrase *catch the bus* is also isolated from other metaphoric terms in a single interpretative repertoire, while the metaphoric acronym *ctb*, as discussed above, co-occurs with *journey* at some max_df densities explored, but not others.

For the cancer.net investigation, metaphoric terms were explored further where they were found to be metaphoric in at least 90% of instances. For suicide.net, this represents the terms *journey*, *ctb*, *battle*, and *struggle*, and the metaphoric phrase *catch the bus*.

However, since the noun *fight*, although it is only used metaphorically in 75% of instances, is present in a metaphoric sense in an interpretative repertoire, *fight* is also considered in subsequent analysis, alongside the analytic insight that it is used literally in 25% of instances.

Although there is a difference in density for the metaphors *ctb* and *catch the bus* between suicide.net.20 and suicide.net.LT20, this is likely to be due to familiarity and repetition, which may lead to higher use of the acronym, rather than because they do different psychological work. For this reason, *ctb* and *catch the bus* are considered as the single metaphor theme *ctb*. The following table summarises the finalised metaphor themes and their constituent terms.

Table 6-18 metaphoric terms for subsequent investigation

metaphor theme	constituent terms
ctb	ctb, catch the bus
journey	journey
battle	battle
fight	fight
struggle	struggle

6.6 Metaphoremes and subject positions

In this section noun phrases relating to the metaphor themes are investigated as a way of identifying typical metaphor use, and metaphoremes, where a metaphoreme, which may

be part of a wider metaphor theme, is a semantic and pragmatic unit of analysis that combines "the linguistic, the cognitive, the affective, and the socio-cultural" (Cameron and Deignan, 2006, p. 686) in a particular speech community at a particular time.

The section starts with graph-based ranking of all noun phrases in the suicide.net.20 data, to provide an overview of the dominant concerns of the community and consider if any of these are represented by metaphor. Subsequently a comparison is made of noun phrases relating to each metaphor theme, between suicide.net.20 and suicide.net.LT20. For each participant group posts were extracted that contain the metaphor theme under investigation, then all noun phrases were extracted from those posts. To make the results more comparable between the groups, the prevalence of each metaphoric phrase as a proportion of all included metaphoric phrases was calculated.

In the first analysis for each metaphor theme, in order to consolidate similar noun phrases, only nouns were retained from the phrases extracted, for example the noun phrases *my ctb method* and *your ctb method* were counted as an instance of the single phrase *ctb method*. In addition, instances of the metaphoric term used alone, for example the single term *journey*, were excluded. In the subsequent analysis for each metaphor theme the noun phrases were retained as a whole, for example the noun phrases *my ctb method* and *your ctb method* were retained as separate phrases. Retaining adjectives, pronouns, articles, and so on, provides information about subject positions used in conjunction with the metaphor theme, as well as particular characterisations of the theme.

To focus on community interpretative repertoires, not idiosyncratic language, phrases are only listed in the results tables where they constitute at least 2% of all relevant phrases and occur at least twice, while non-metaphoric phrases meeting these criteria are crossed out.

6.6.1 Graph-based ranking of the suicide.net data

In an application of graph-based ranking to the suicide.net.20 data noun lemmas were used to calculate the rank of individual noun phrases because, as discussed in chapter two, the focus in the current study is on commonly used metaphoric terms in this data as they are instantiated as nouns. In order to concentrate similar phrases based on the same underlying metaphor, each noun phrase was contracted to only its nominal components. Because of the still very wide range of individual phrases in comparison to cancer.net, phrases were included that constitute at least 1% of all the extracted phrases. In addition, the analysis was run on a sample of 10,000 posts. This produced the single ranked phrase *survival instinct*, which occurred 76 times in the sample, and is present 2,717 times in suicide.net as a whole. The phrase *survival instinct* was also shown in the key term analysis (section 6.3.2) to be the highest ranked key term, occurring over 1,000 times as often in suicide.net as it does in the comparison general web corpus.

6.6.2 journey phrases

For this analysis posts were included in which there is at least one instance of the *journey* metaphor theme, and from those posts noun phrases were extracted that contain the noun *journey*.

6.6.2.1 journey noun phrases, nouns only

In the first results table below, in which only their constituent nouns are used to represent noun phrases, for more established participants (suicide.net.20, on the left of the table) the phrase *journey friend* was found to be the most prevalent, where *journey friend* represents phrases such as *enjoy the journey friend* and *have a safe journey friend*. A concordance analysis found that the association of the noun *friend* after *journey* is only used where *journey* is metaphoric, never when it is literal. The phrase *journey brother*,

which is the third most frequent phrase for more established participants, was found to be used in a similar way to *journey friend*, for example *good luck on your journey brother*.

The second most common *journey* phrase for more established participants was found to be *sn journey*, which is a metonymy for suicide using sodium nitrite, where sodium nitrite is a sought-after suicide method. The sparse number of *journey* phrases extracted for the less established participants (suicide.net.LT20, on the right of the table) suggests there are no *journey* metaphoremes for that group; rather, there are a wide range of diverse *journey* phrases that do not share similar noun components. For the less established participants, then, *journey* is not a typical metaphor. In addition, for more established participants the longer timeframe phrase *life journey* is present, while for less established participants the shorter-term timeframe *year journey* is present.

Table 6-19 journey noun phrases, nouns only

	suicide.net.20	count	% of phrases	suicide.net.LT20	count	% of phrases
1.	journey friend	46	21	*year journey	2	18
2.	sn journey	41	19	ctb journey	2	18
3.	journey brother	8	4	journey friend	2	18
4.	journey end	6	3			
5.	*life journey	5	2			
6.	hour journey	4	2			
7.	journey luck	4	2			
8.	journey note	4	2			
9.	journey lot	4	2			

6.6.2.2 journey noun phrases, all parts of speech

In the following results table, *journey* noun phrases were collected based on all parts of speech, not just nouns. In this case the most common *journey* phrase for both participant groups was found to be *your journey*. The phrase *your journey*, which represents over 20% of all *journey* phrases for both groups, is a metaphoreme in this context that is used to convey support for the suicide of another participant, a value that is perhaps the most

antithetical to the stance towards suicide of wider society. This supports the previous insight that metaphoric *journey* is typically used to address the experience of another participant, which in the context of suicide.net typically means their *journey to catch the bus*.

The most notable difference between the two groups is that the phrase *my journey* accounts for 2% of suicide.net.20 *journey* phrases, and 7% of suicide.net.LT20 *journey* phrases, again reinforcing the insight that more recent participants are more focused on their own experience and have not yet extended their focus outwards to the community; they perhaps do not yet feel a sense of membership of the community.

Table 6-20 journey noun phrases, all parts of speech

	suicide.net.20	count	% of phrases	suicide.net.LT20	count	% of phrases
1.	your journey	559	22	your journey	045	24
2.	a peaceful journey	353	14	a peaceful journey	024	13
3.	a safe journey	217	8	my journey	013	7
4.	journey	136	5	the journey	010	5
5.	the journey	127	5	safe journey	009	5
6.	safe journey	110	4	a safe journey	007	4
7.	my journey	062	2	journey	005	3
8.	a journey	058	2	this journey	005	3
9.	a safe and peaceful journey	053	2	a good journey	004	2
10.	this journey	046	2	a journey	003	2

The following examples from a concordance analysis of the phrase *my journey* show that it is used more in accordance with the conventional LIFE IS A JOURNEY conceptual metaphor, i.e. for considering a single life as a whole as a journey, rather than as an extension of the *ctb* metaphor with its shorter-term focus, for which related *journey* on suicide.net is typically the act of suicide.

1. this made me the school clown and started my journey of being worth less than others

2. some would argue it's the journey that matters but so far, my journey has been a total shit show
3. my journey on this planet is coming to an end
4. i also feel like suicide is part of my journey
5. my story is my own, my journey is as unique as all of yours
6. they are part of me and my journey in this hell called life

6.6.3 ctb phrases

For this analysis posts were included in which there is at least one instance of the *ctb* metaphor theme, and from these noun phrases were extracted that contain the term *ctb*. It was shown previously that the metaphoric term *ctb* is dominant in both more established and less established participant groups, and is a metaphoreme, in that it has become a shorthand way of describing particular meaning and related psychological work based on the metaphoric phrase *catch the bus*.

6.6.3.1 ctb noun phrases, nouns only

In the first analysis in the current section consideration is given to nouns that occur alongside *ctb*. It can be seen from the first results table below that use of *ctb* is similar in both participant groups, with phrases that are common to both groups having similar densities. The single time reference for each group is *ctb day*.

Table 6-21 ctb noun phrases, nouns only

	suicide.net.20	count	% of phrases		suicide.net.LT20	count	% of phrases
1.	ctb method	587	14		ctb method	56	19
2.	ctb date	346	8		ctb date	24	8
3.	ctb attempt	323	8		ctb plan	18	6
4.	ctb plan	280	7		ctb attempt	14	5
5.	ctb partner	110	3		*ctb day	8	3
6.	ctb thought	96	2		ctb partner	8	3

7.	*ctb day	72	2	ctb option	6	2
8.				ctb amp way	6	2
9.				ctb time	6	2
10.				ctb ing	6	2
11.				ctb por dejar	6	2

6.6.3.2 ctb noun phrases, all parts of speech

For the analysis in this section, *ctb* noun phrases were collected based on all their component parts of speech, not just nouns. It can be seen that again the metaphoric phrases that include *ctb* are similar in suicide.net.20 and suicide.net.LT20 and have a similar density. The self references *my ctb*, *my ctb date*, and *my ctb plan* also have similar densities for each participant group. In suicide.net.20 the phrase *your ctb* is also present, while in suicide.net.LT20 no references to the other are present. This supports previous insight that even though all participants use *ctb* to discuss various aspects of suicide, the orientation of more established participants is also towards the other, while the orientation of newer participants is predominantly towards the self.

Table 6-22 ctb noun phrases, all parts of speech

suicide.net.20			% of phrases	suicide.net.LT20		
	count				count	% of phrases
1.	my ctb	463	8	my ctb	30	7
2.	my ctb date	194	3	ctb method	18	4
3.	your ctb	188	3	my ctb date	14	3
4.	my ctb plan	134	2	a ctb method	12	3
5.	ctb method	108	2	that ctb	10	2
6.	the ctb	97	2	a ctb	9	2
7.				my ctb plan	8	2
8.				what ctb	7	2

6.6.4 fight phrases

For this analysis posts were included in which there is at least one instance of the *fight* metaphor theme, and from these noun phrases were extracted that contain the noun lemma *fight*.

6.6.4.1 fight noun phrases, nouns only

For this analysis, in which only nouns were used to represent their corresponding noun phrases, for both groups of participants the most prevalent phrase is literal; for more established participants it refers to 'Fight Club', the movie. For the less established participants there are insufficient phrases to support an analysis, except to say that there is no dominant *fight* phrase for that group, while for more established participants there are two references to a fight or flight response. These results suggest that there are no *fight* metaphoremes in this context.

Table 6-23 fight noun phrases, nouns only

suicide.net.20			suicide.net.LT20		
	count	% of phrases		count	% of phrases
1. fight club	22	8	fight irl	2	40
2. fight flight mode	16	6			
3. huge fight	15	5			
4. good fight	13	5			
5. fight mode	12	4			
6. more fight	9	3			
7. much fight	9	3			
8. big fight	8	3			
9. hellish fight flight freeze condition	6	2			
10. physical fight	6	2			
11. increase fight flight response	6	2			

6.6.4.2 fight noun phrases, all parts of speech

For the results listed in the following table *fight* noun phrases were collected based on all parts of speech. It can be seen that in this case the most common *fight* phrases are similar for both groups, with the noun phrase *a fight* the most prevalent *fight* noun phrase for both groups. The first-person singular reference *my fight* is present in the group of more established participants, but not the group of more recent participants. However, there are

far fewer occurrences of *fight* to investigate, and as with *my journey* discussed above, *my fight* in suicide.net.20 occurs as 2% of all *fight* phrases.

Table 6-24 *fight* noun phrases, all parts of speech

suicide.net.20	count	% of phrases	suicide.net.LT20	count	% of phrases
1. a fight	165	22	a fight	22	25
2. the fight	105	14	the fight	10	11
3. a fighter	027	4	no fight	4	4
4. this fight	023	3	the fights	2	2
5. my fight	015	2	that fight	2	2
6. a fighting chance	013	2	a physical fight	2	2
7. a huge fight	012	2	a big fight	2	2
8. no fight	012	2			

6.6.5 battle phrases

For this analysis posts were included in which there is at least one instance of the *battle* metaphor theme, and from these noun phrases were extracted that contain the noun lemma *battle*.

6.6.5.1 battle noun phrases, nouns only

It can be seen from the results table below that when only nouns are used to represent their corresponding *battle* noun phrases there are a larger number of *battle* phrases for the more established group, with *custody battle* the most prevalent phrase. As with *fight*, there does not seem to be a dominant suicide.net *battle* metaphoreme, although there are multiple references to 'Battle Royale', a genre based on the dystopian novel of the same name (Takami, 2009), in which a totalitarian state forces high school students to battle each other to the death.

Table 6-25 battle noun phrases, nouns only

	suicide.net.20	count	% of phrases	suicide.net.LT20	count	% of phrases
1.	custody battle	9	9	battle royale style survival	6	50
2.	battle field	6	6	custody battle	2	17
3.	battle sign	6	6			
4.	battle royale style office politic culture	6	6			
5.	*year battle	5	5			
6.	battle rifle machine gun	5	5			
7.	court battle	4	4			
8.	battle depression	4	4			
9.	battle trench	4	4			
10.	battle soldier	3	3			
11.	hill battle	3	3			
12.	battle disorder	3	3			
13.	mother battle stomach cancer	3	3			
14.	battle network series	3	3			
15.	battle station	3	3			
16.	battle scar	3	3			
17.	battle struggle	3	3			
18.	film battle royale	2	2			
19.	battle bullshit	2	2			
20.	battle royale	2	2			
21.	*month battle	2	2			
22.	battle trailer	2	2			
23.	battle mage	2	2			
24.	battle phrase	2	2			
25.	battle addiction	2	2			
26.	battle buddy	2	2			
27.	battle ax	2	2			
28.	battle hymn	2	2			

6.6.5.2 battle noun phrases, all parts of speech

When all parts of speech were used to represent *battle* noun phrases, the pronoun-based phrases *my battle* and *your battle* were found to be present in the group of more established participants, but not the group of more recent participants. However, there is a

relatively low density of *battle* phrases, which in addition are only listed where they occur at least twice and constitute at least 2% of all selected phrases. There is insufficient data for the suicide.net.LT20 group for conclusions to be drawn. It is notable that in comparison to cancer.net, in which *battle* is predominantly used to talk about periods of time, no time phrases are present in these results.

Table 6-26 battle noun phrases, all parts of speech

	suicide.net.20	count	% of phrases	suicide.net.LT20	count	% of phrases
1.	the battle	70	14	the battle	8	16
2.	a battle	57	11	a battle	8	16
3.	this battle	32	6	some battle royale style social - survival	6	12
4.	a losing battle	18	4	a losing battle	3	6
5.	a constant battle	15	3	this battle	2	4
6.	an uphill battle	14	3	constant battle	2	4
7.	my battle	13	3			
8.	your battle	10	2			

6.6.6 struggle phrases

For this analysis posts were included in which there is at least one instance of the *struggle* metaphor theme, and from these noun phrases were extracted that contain the noun lemma *struggle*.

6.6.6.1 struggle noun phrases, nouns only

It can be seen from the results table below that, as with the same analysis for *battle*, above, there are a wide range of *struggle* phrases when only nouns from those phrases are used to represent them, with *health struggle* the most prevalent phrase. As for *fight*, and *battle*, there does not seem to be a dominant *struggle* metaphoreme. And while *struggle* is comparatively prevalent in suicide.net compared to the other metaphor themes, for the less established group none of the *struggle* noun phrases meet the conditions for inclusion

in the results table: there are too many disparate terms. This is notable given that *struggle* occurs approximately twice as often as *fight* for both groups. The diversity of *struggle* phrases is also consistent with the fact that *struggle* was found to be present in two interpretative repertoires, whereas the other primary metaphoric terms under investigation were each present in only one interpretative repertoire, suggesting that *struggle* has more diverse uses.

Table 6-27 struggle noun phrases, nouns only

	suicide.net.20	count	% of phrases	suicide.net.LT20	count	% of phrases
1.	health struggle	15	13			
2.	life struggle	9	8			
3.	power struggle	8	7			
4.	people struggle	7	6			
5.	struggle mark	6	5			
6.	person struggle	4	3			
7.	struggle death throe	4	3			
8.	*lifetime struggle	4	3			
9.	struggle being	3	3			
10.	struggle injury	3	3			
11.	struggle man	2	2			
12.	*struggle day	2	2			
13.	family power struggle	2	2			
14.	son struggle	2	2			
15.	*year struggle	2	2			
16.	term struggle	2	2			
17.	poverty struggle	2	2			
18.	struggle increase	2	2			
19.	struggle imaginable	2	2			
20.	struggle opening	2	2			
21.	struggle dude	2	2			

6.6.6.2 struggle noun phrases, all parts of speech

In the following table, *struggle* noun phrases are collected based on all parts of speech. It can be seen that the most common *struggle* phrase for both participant groups is *the struggle*, followed by *a struggle*. Although not many *struggle* noun phrases meet the criteria for inclusion in the results table when all parts of speech are included, it is notable that *your struggle*, referring to the other, is more than twice as prevalent for the more established group in comparison to the less established group (10% in comparison to 4%), while the density of *my struggle*, referring to the self, is similar for both groups.

Table 6-28 struggle noun phrases, all parts of speech

suicide.net.20			suicide.net.LT20		
	count	% of phrases		count	% of phrases
1. the struggle	153	14	the struggle	22	18
2. a struggle	132	12	a struggle	19	15
3. your struggle	116	10	my struggle	8	6
4. my struggle	73	7	a constant struggle	5	4
5. their struggle	32	3	your struggle	5	4
6. our struggle	22	2	*the daily struggle	4	3
7. his struggle	18	2	*a daily struggle	3	2
8.			our struggle	3	2
9.			personal struggle	2	2
10.			such a struggle	2	2
11.			the constant struggle	2	2
12.			my big struggle	2	2

6.6.7 Section summary: metaphoremes and subject positions

The phrases *ctb* and *your journey* have been identified as metaphoremes, and possibly related, in the suicide.net context. While *journey* in this context has been found to be predominantly used to wish someone else well when they *ctb*, the phrase *my journey* in comparison was found to be typically used in a manner more consistent with the hegemonic LIFE IS A JOURNEY conceptual metaphor, to frame a single life as a journey in a longer-term sense than *ctb* and typical suicide.net use of the *journey* theme

entails. The phrase *my journey* was found to be over three times more prevalent in the suicide.net.LT20 discourse: in the data of more recent participants the *journey* metaphor has perhaps not yet been recontextualised, where recontextualisation is the re-use of concepts in different contexts to fulfil different needs (Linell, 2009; Semino, Deignan and Littlemore, 2013).

The analysis of *fight*, *battle*, and *struggle*, each of which was found to be used more often by less established than by more established participants, showed that there is a wide range of noun phrases associated with these terms, such that there is no sense of them having a typical use in suicide.net, and no obvious metaphoremes. In terms of subject positions, the other reference *your* was associated with the more established group for *battle* and *struggle* but not for *fight*; for *fight*, the self reference *my fight* was associated with the more established but not the less established group. The results for *fight* may also reflect the fact that *fight* instantiated as a noun is metaphoric in only 75% of uses.

In terms of potentially more deliberate metaphor use, unlike *journey*, *fight*, and *battle*, *struggle* does not appear to invoke a particular source domain strongly. For example, the master metaphor list (Lakoff, Espenson and Schwartz, 1991) categorises *struggle* as a mental event, in comparison to *journey* which it categorises as both an event structure mapping, and as an emotion mapping (for example, LOVE IS A JOURNEY), while *fight* and *battle* are both categorised as both a mental event and an event structure mapping. In addition the ratio of use of *struggle* in suicide.net.20 in comparison to suicide.net.LT20 is 0.84, which is by far the lowest ratio, such that it can not considered to be a characteristic metaphor of this community, but rather a more general metaphor that is perhaps displaced by actual community metaphor in the discourse of more established suicide.net participants. While as has been discussed in chapter one, pervasive, more instinctive,

metaphors are recognised as having great importance and impact on language and cognition, they are not the focus of the current study, which seeks metaphor themes that characterise a particular community, and for which they have a particular communicative and rhetorical meaning. For these reasons the *struggle* metaphor theme is not considered in future analysis in this study.

6.7 Metaphor and general language style

In this section an investigation is made into whether each metaphor theme is surrounded by a particular distinct language style, such that it may be said to perform particular social and psychological work. As with the same investigation for cancer.net, which is discussed in detail in chapter four, a supervised learning algorithm was trained to predict the presence or absence of each metaphor theme in any post in suicide.net, based on the language style of that post. Two supervised learning methods were used and compared. A text-based method supplied the text of each post to the algorithm, while a count-based method supplied the counts of 143 different linguistic variables (section 4.4.1) from each post. For both methods the relevant metaphoric terms were first removed from the data supplied to the algorithm.

The model produced by the algorithm was subsequently used to predict the presence or absence of each metaphor theme in a set of test posts. The accuracy of the model's predictions is used to support the organisation of metaphoric terms into themes, and to support the concept that different metaphors are surrounded by particular language styles, and that they each therefore do specific social and psychological work, as was found to be the case for cancer.net.

6.7.1 Text-based method

For the text-based supervised learning method, the whole text of each post was supplied to the algorithm, with the metaphoric terms under investigation first having been removed from that text. The results are summarised in the table below and ranked by weighted average F1 score, which is a standard measure of the accuracy of a model. It can be seen that the model was most accurate in predicting the *journey* condition (F1=0.81), and least accurate for predicting the *ctb* condition (F1=0.70), although all the results are credible compared to similar investigations (Shutova, 2015; Leong *et al.*, 2020). The results for each metaphoric condition are discussed in more detail in subsequent sections.

Table 6-29 weighted average F1 score of text-based supervised model in different metaphoric conditions

metaphoric condition	weighted average F1	test posts in condition 0	test posts in condition 1
journey	0.81	993	881
fight	0.73	234	244
battle	0.72	195	183
ctb	0.70	3893	3947

6.7.1.1 journey condition

The classification report and confusion matrix below show that for the *journey* metaphor there is an imbalance between precision and recall, with the model more likely to correctly predict posts that do not contain the *journey* theme than posts that do. This is visible in the subsequent scatter plot, which also shows a concentration of posts very confidently correctly predicted as containing the *journey* theme. This suggests a particular subset of this theme that has not been identified as a separate usage: perhaps the *journey* theme as identified here is a composite of multiple themes doing distinct work.

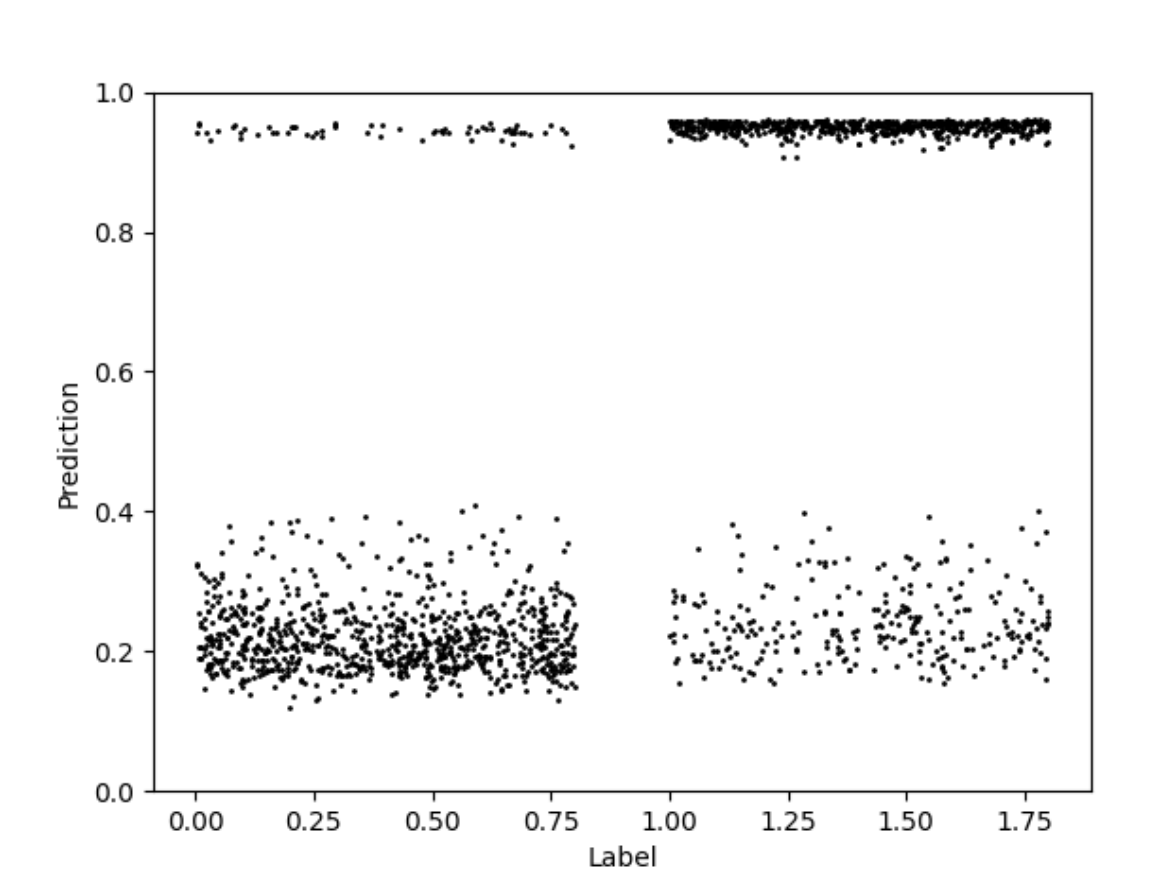
Table 6-30 classification report: journey

	precision	recall	f1-score	support
0	0.76	0.93	0.84	933
1	0.90	0.69	0.78	881
accuracy			0.81	1814
macro avg	0.83	0.81	0.81	1814
weighted avg	0.83	0.81	0.81	1814

Table 6-31 confusion matrix: journey

confusion matrix		predicted 0	predicted 1
true 0		869	64
true 1		272	609

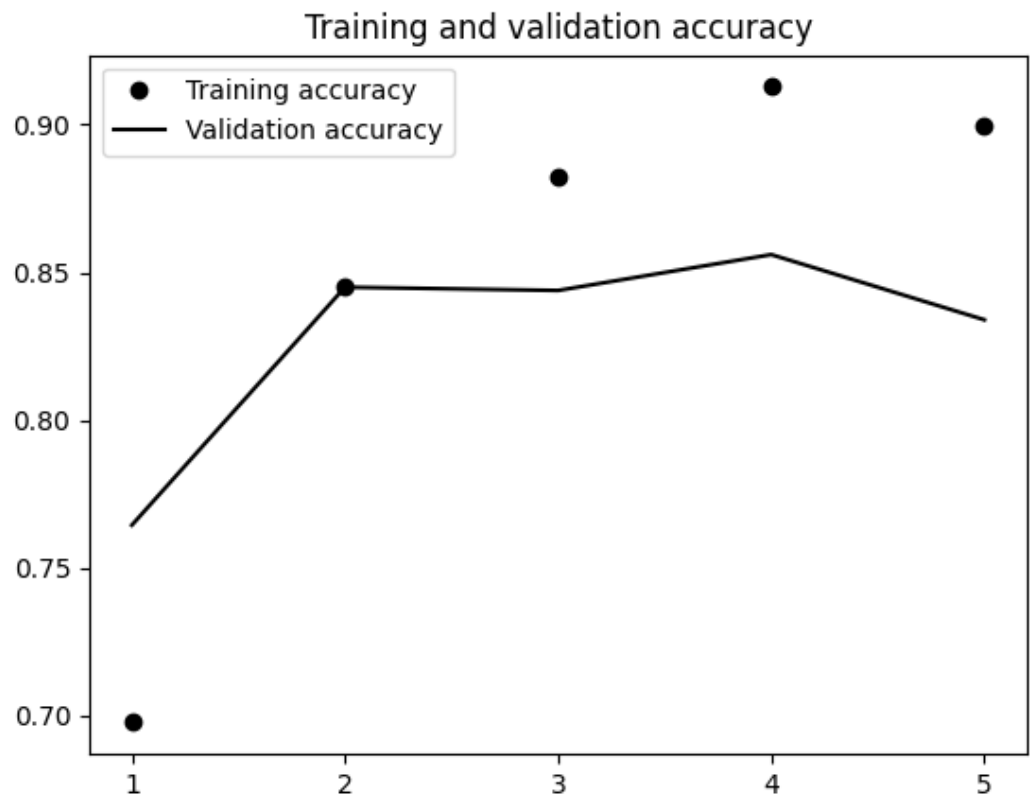
Figure 6-2 scatter plot of predictions: journey



The following plot of training and validation accuracy over five epochs confirms that around five epochs is appropriate for this analysis. Training accuracy continues to

increase in epochs three and four while validation accuracy levels off, then validation accuracy falls more steeply at epoch five, suggesting that training over more than five epochs could lead to overfitting of the model to the training data.

Figure 6-3 training and validation accuracy: journey



6.7.1.2 ctb condition

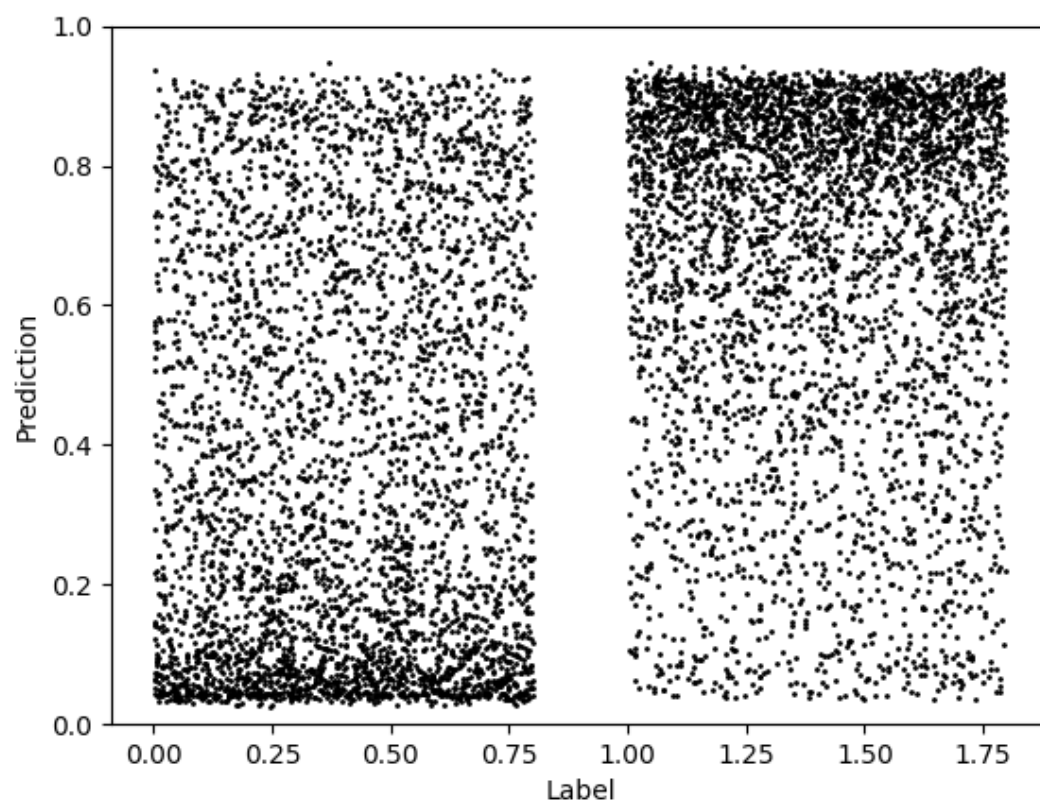
The classification report and confusion matrix below show that for the *ctb* metaphor precision and recall are more balanced than they were for the *journey* theme. This again is visible in the subsequent scatter plot, which shows a similar distribution for each class.

Table 6-32 classification report: *ctb*

	precision	recall	f1-score	support
0	0.72	0.64	0.68	3893
1	0.68	0.76	0.71	3947
accuracy			0.70	7840
macro avg	0.70	0.70	0.70	7840
weighted avg	0.70	0.70	0.70	7840

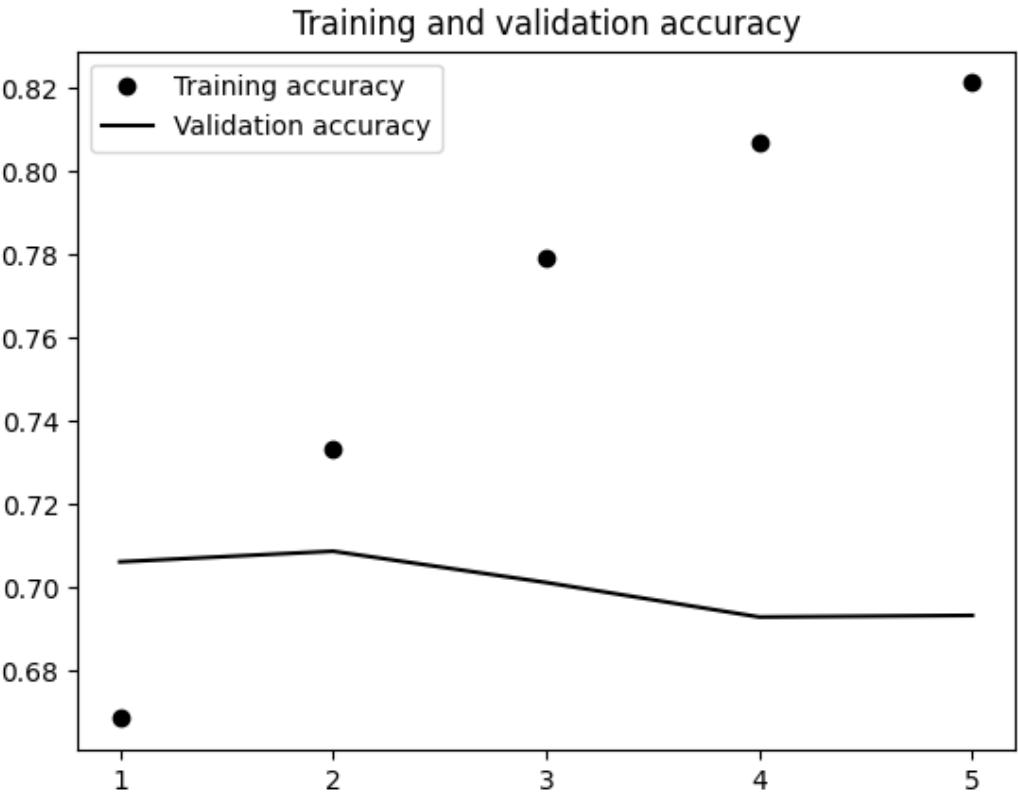
Table 6-33 confusion matrix: *ctb*

confusion matrix		predicted 0	predicted 1
true 0		2476	1417
true 1		964	2983

Figure 6-4 scatter plot of predictions: *ctb*

As for the *journey* theme above, the plot of training and validation accuracy below shows that to avoid overfitting, around five epochs is right for the *ctb* theme.

Figure 6-5 training and validation accuracy: *ctb*



6.7.1.3 fight condition

The classification report and confusion matrix below show that recall and precision are relatively less balanced for the *fight* theme than for the *ctb* theme. The model is more likely to identify correctly posts labelled as not containing the *fight* metaphor theme, than posts labelled as containing the theme.

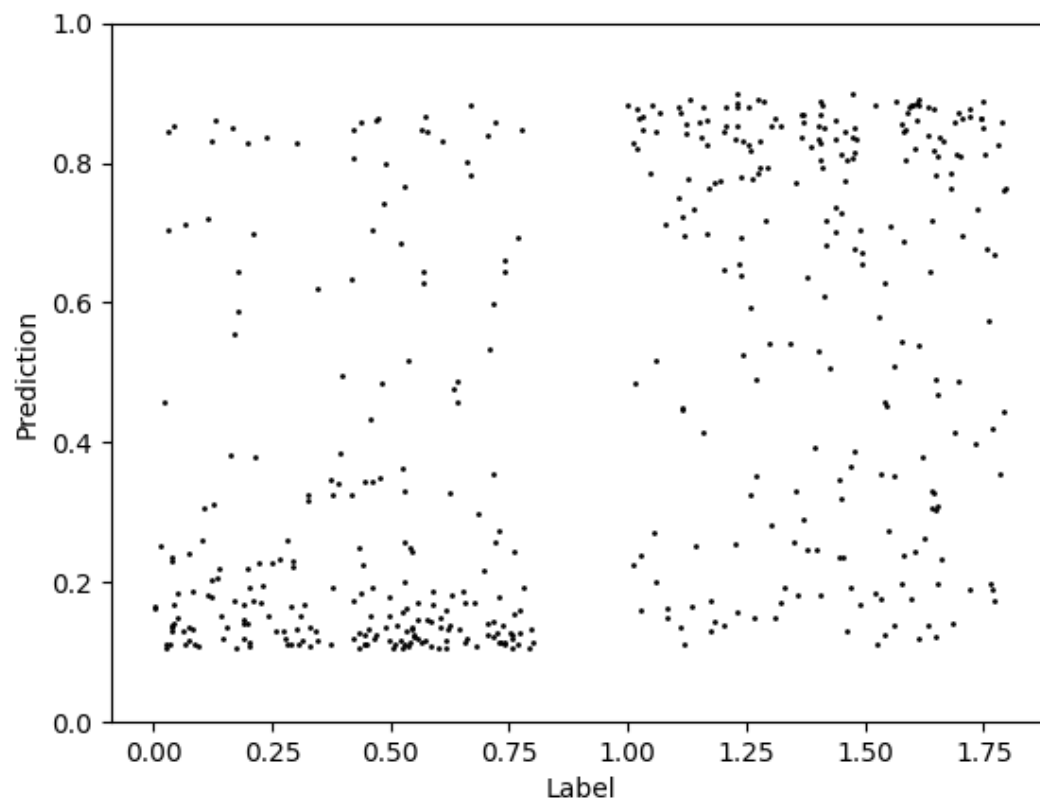
Table 6-34 classification report: fight

	precision	recall	f1-score	support
0	0.69	0.81	0.74	234
1	0.78	0.65	0.71	244
accuracy			0.73	478
macro avg	0.73	0.73	0.73	478
weighted avg	0.74	0.73	0.73	478

Table 6-35 confusion matrix: fight

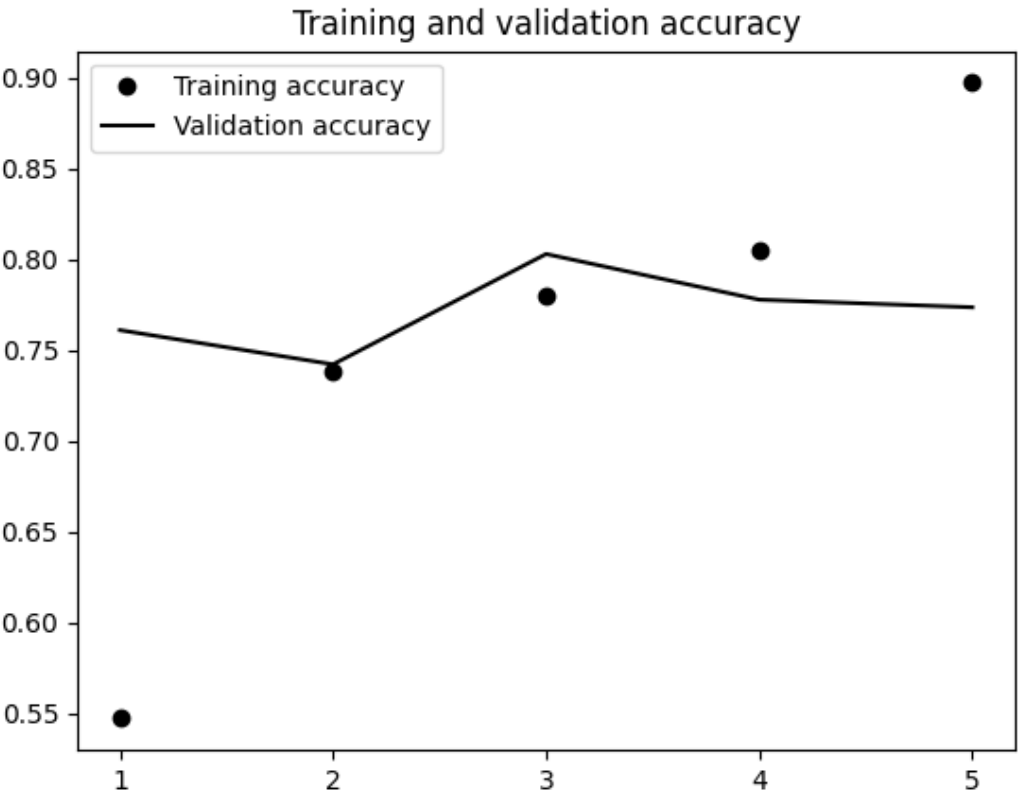
confusion matrix		predicted 0	predicted 1
true 0		189	45
true 1		85	159

Figure 6-6 scatter plot of predictions: fight



As for the *journey* and *ctb* themes above, the plot of training and validation accuracy over five epochs below shows that to avoid overfitting, around five epochs is right for the *fight* theme.

Figure 6-7 training and validation accuracy: fight



6.7.1.4 battle condition

The classification report and confusion matrix below show that as for the *journey* metaphor precision and recall are relatively unbalanced for the *battle* theme. This again is visible in the subsequent scatter plot, which shows a concentration of confidently and correctly predicted metaphoric posts (condition 1), while there is a more uniform spread of predictions over the whole range for posts labelled as non-metaphoric. However, there is a small number of posts to support this investigation.

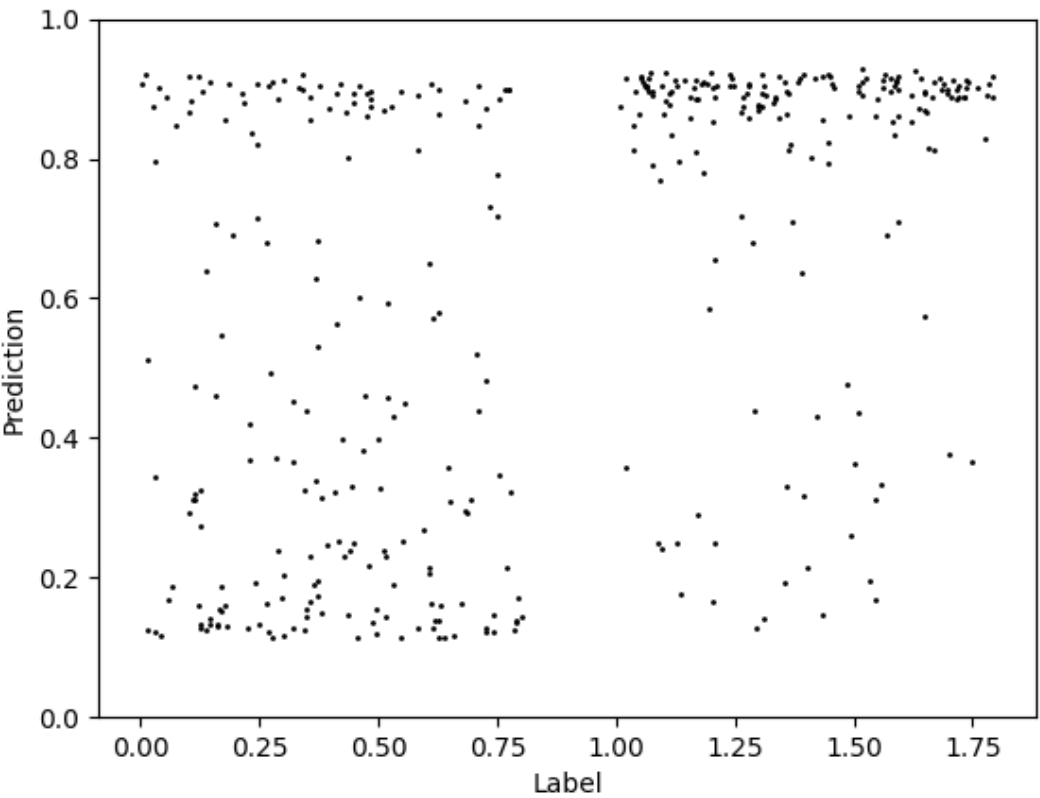
Table 6-36 classification report: battle

	precision	recall	f1-score	support
0	0.81	0.59	0.69	195
1	0.66	0.85	0.75	183
accuracy			0.72	378
macro avg	0.74	0.72	0.72	378
weighted avg	0.74	0.72	0.72	378

Table 6-37 confusion matrix: battle

confusion matrix		predicted 0	predicted 1
true 0		116	79
true 1		27	156

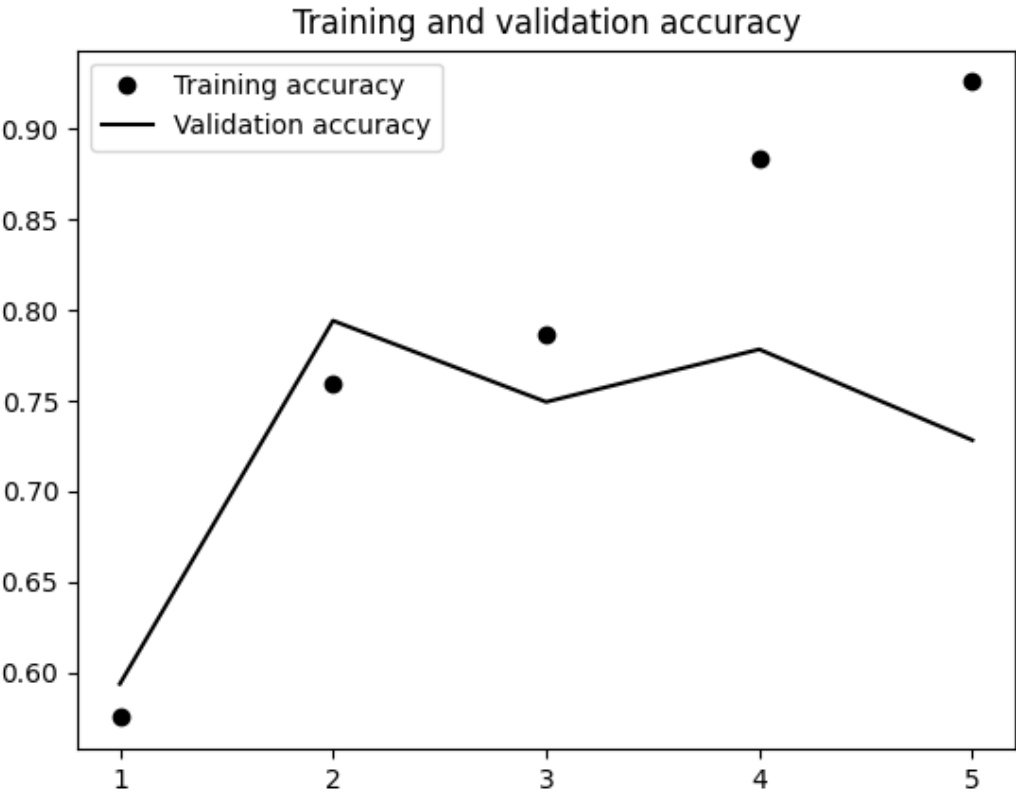
Figure 6-8 scatter plot of predictions: battle



As for the other metaphor themes investigated above, the plot of training and validation accuracy over five epochs below shows that to avoid overfitting, around five epochs is

right for the *battle* theme, or if anything fewer than five: if the number of epochs is greater than around five, accuracy of the model is likely to be adversely affected.

Figure 6-9 training and validation accuracy: battle



6.7.2 Count-based method

In this count-based supervised learning method, the counts of 143 different linguistic variables were used (section 4.4.1): the actual post text was not supplied to the supervised algorithm. For each metaphoric condition the metaphoric terms under investigation were removed from the text before the counts of linguistic variables were made. In addition the counts of the linguistic variables were converted to proportions of the word count of the complete text so that the length of a post was not dominant in providing information to the algorithm about language style, since, as is shown in a subsequent analysis, word count is much higher in posts in which the metaphor themes are used than in posts in

which they are not used. Five-fold stratified cross-validation was used to split the data into sample and test sets, with mean weighted average F1 used as the measure of accuracy of the model.

It can be seen from the results table below that this count-based method predicts each metaphoric condition more accurately than does the text-based method considered previously. The *journey* condition was again the most accurately predicted (F1=0.92); *battle* was the least well predicted (F1= 0.83). In addition, precision and recall are relatively well balanced for each condition. All the results are credible compared to similar investigations, and this supports the organisation of the metaphor themes, and demonstrates that a distinctive language style surrounds use of each theme.

Table 6-38 accuracy of count-based model in different metaphoric conditions

metaphoric condition	weighted average F1	precision	recall	test posts
journey	0.92	0.91	0.93	1088
ctb	0.90	0.87	0.94	4702
fight	0.87	0.85	0.90	286
battle	0.83	0.90	0.88	227

6.7.3 Section summary: metaphor and general language style

The high accuracy scores obtained when two supervised learning algorithms using data represented in different ways were trained to predict the presence or absence of a particular metaphor theme in a post show that the different metaphor themes found in this community are surrounded by different language styles, and hence may be considered to perform specific, and different, social and psychological work in this context. As with the same analysis for cancer.net, the count-based method, which is based on counting 143 different linguistic variables in each post, was more accurate than the text-based method, in which post texts were supplied to the algorithm.

6.8 Metaphor and specific language style

In this section the difference in language style between different metaphoric conditions is considered in more detail. As with the same analysis carried out for cancer.net, which is discussed in more detail in chapter four, for each metaphor theme separately the density of 143 linguistic variables was compared between posts that contain that metaphor theme (corpus F, the corpus of focus), and posts that do not contain the theme (corpus R, the reference corpus). Variables for which there is a significant difference in density between the two corpora are listed in the results tables, in this case where the log likelihood of the difference is greater than 15.13 ($p < 0.0001$), and Cohen's d is greater than 4, where Cohen's d values greater than 2 and less than 5 represent a small effect, and values greater than 5 represent a medium effect. The score was calculated on the same basis used for keywords and key terms above, and the smoothing parameter N was set to 1. In addition, based on specific differences in language style that have been related to state of mind, the standard deviation of pronoun types used was calculated for each condition, as well as the ratio of positive to negative emotion words. As with the supervised learning analysis in the previous section, the metaphoric terms were removed from the text before linguistic variables were counted.

Since the focus in this investigation is on personal pronouns and emotion words as a particular site of psychological work, where such terms are increased in corpus F they are marked in bold in the results tables; where they are decreased in corpus F they are underlined. However, since this is a linguistics investigation, all significantly different linguistic variables are listed.

Because the very diverse data of suicide.net is often structured as an ongoing chat with short replies between participants, there are many posts with a small number of words,

with some posts containing only punctuation, while others are simple phrases that do not constitute a complete sentence. In comparison the posts in which metaphor is present tend to be longer. For this reason, in order to focus on linguistic differences in more comparable text, posts were included in this analysis where they are ten or more tokens in length.

6.8.1 journey

In this analysis the data was split based on the presence or absence of the *journey* metaphor theme. Corpus F consists of all the posts that have at least one instance of the *journey* theme. Corpus R consists of all the posts that do not contain at least one instance of the *journey* metaphor theme.

It can be seen from the results table below that there are nearly twice as many positive as negative words in corpus F, while in corpus R there is a similar density of positive and negative emotion words. In addition, a wider range of personal pronouns is used in corpus F in comparison to corpus R.

Table 6-39 density of journey in suicide.net

	posts	tokens	average post length (tokens)	standard deviation of pronouns	ratio positive to negative emotion words
Corpus F journey>0	2,148 (0.6%)	207,554	97	2.17	1.80
Corpus R journey=0	354,791	24,783,525	70	1.29	0.99

The table below lists the eight linguistic variables that were found to be significantly different in density between posts in which *journey* is and is not used, ranked according to the proportion of change between corpus F and corpus R. It can be seen that the positive emotion variables *Joy*, *Surprise*, *Trust*, *Positive*, and *Anticipation*, are all significantly more prevalent where *journey* is present in posts. The pronoun *your* is also

significantly more used, reflecting the dominant use of *journey* as the metaphoreme *your journey*, and there are no significantly decreased variables associated with use of *journey*. This result is in contrast to use of the *journey* theme in cancer.net which was associated with a much more complex language style including 61 significantly different linguistic variables.

Table 6-40 most changed linguistic variables between posts that do and do not contain journey

	variable	freq F	freq R		% change	LL	Cohen's d	score
1.	dative	5184	1640	+	216.03	992	0.65	3.16
2.	your	9602	4357	+	120.39	962	0.49	2.20
3.	Joy	37903	19253	+	96.87	2885	0.64	1.97
4.	you	31765	17405	+	82.50	1951	0.57	1.83
5.	Surprise	23329	14339	+	62.70	972	0.48	1.63
6.	Trust	42558	27657	+	53.88	1415	0.47	1.54
7.	Positive	64981	43800	+	48.36	1830	0.44	1.48
8.	Anticipation	42110	28604	+	47.21	1143	0.43	1.47

6.8.2 *ctb*

In this analysis the data was split based on the presence or absence of the *ctb* metaphor theme. Corpus F consists of all posts that have at least one instance of the *ctb* theme, and corpus R consists of all posts that do not contain at least one instance of the *ctb* theme. It can be seen from the first results table below that there is little difference in the ratio of positive to negative words between corpus F and corpus R, but there is a wider range of personal pronouns in corpus F in comparison to corpus R.

Table 6-41 density of *ctb* in suicide.net

	posts	tokens	average post length (tokens)	standard deviation of pronouns	ratio positive to negative emotion words
Corpus F <i>ctb</i> >0	11,512 (3.23%)	1,514,242	132	2.40	1.05
Corpus R <i>ctb</i> =0	345,427	21,149,769	68	1.24	0.99

The table below lists the 38 variables that were found to be significantly different in density between posts in which *ctb* is and is not used, ranked according to the proportion of change between corpus F and corpus R. It can be seen that the most increased variable in corpus F is *PRP\$*, a composite variable of possessive pronouns. The first-person singular pronoun *i* is also increased, and the positive emotion variables *Anticipation* and *Positive* are also significantly increased. No key variables relating to language style were found to be significantly decreased.

Table 6-42 most changed linguistic variables between posts that do and do not contain *ctb*

	variable	freq F	freq R		% change	LL	Cohen's d	score
1.	PRP\$	21788	18224	+	19.56	929	0.44	1.20
2.	poss	22043	18726	+	17.71	788	0.43	1.18
3.	CCONJ	41497	37143	+	11.72	697	0.45	1.12
4.	i	48249	43305	+	11.42	772	0.44	1.11
5.	TO	25002	22486	+	11.19	385	0.47	1.11
6.	conj	42217	37994	+	11.11	643	0.41	1.11
7.	VBG	25267	23022	+	9.75	301	0.45	1.10
8.	unique_words	604001	668859	-	9.70	9269	0.59	0.90
9.	Anticipation	31179	28558	+	9.18	332	0.46	1.09
10.	IN	108675	100165	+	8.50	999	0.47	1.08
11.	xcomp	20422	18832	+	8.45	186	0.45	1.08
12.	mark	22643	20904	+	8.32	200	0.44	1.08
13.	ADP	94453	87287	+	8.21	814	0.46	1.08
14.	SCONJ	21113	19516	+	8.18	181	0.44	1.08
15.	prep	82411	76317	+	7.98	673	0.45	1.08
16.	advcl	24079	22384	+	7.57	178	0.44	1.08
17.	PART	43636	40838	+	6.85	266	0.48	1.07

18.	aux	70539	66190	+	6.57	397	0.51	1.07
19.	VB	66297	62351	+	6.33	348	0.50	1.06
20.	Positive	46168	43835	+	5.32	173	0.43	1.05
21.	monosyllabic words	787708	748060	+	5.30	2935	0.50	1.05
22.	PRON	143671	136707	+	5.09	496	0.47	1.05
23.	ADV	80187	76337	+	5.04	271	0.47	1.05
24.	advmod	78219	74672	+	4.75	236	0.47	1.05
25.	pobj	70543	67492	+	4.52	193	0.41	1.05
26.	RB	87386	83712	+	4.39	226	0.47	1.04
27.	VERB	169600	162716	+	4.23	408	0.49	1.04
28.	punctuation	125699	120662	+	4.18	295	0.42	1.04
29.	ccomp	31461	30498	+	3.16	43	0.45	1.03
30.	PRP	106817	103578	+	3.13	143	0.46	1.03
31.	nsubj	116909	120600	-	3.06	162	0.45	0.97
32.	AUX	69962	67993	+	2.90	80	0.49	1.03
33.	MD	22959	22330	+	2.82	25	0.45	1.03
34.	JJ	63815	62309	+	2.42	51	0.41	1.02
35.	ADJ	70036	68495	+	2.25	49	0.42	1.02
36.	DET	81905	80193	+	2.13	52	0.43	1.02
37.	DT	75275	73998	+	1.73	31	0.43	1.02
38.	dobj	58149	57222	+	1.62	21	0.44	1.02

6.8.3 fight

In this analysis the data was split based on the presence or absence of the *fight* metaphor theme. Corpus F consists of all the posts that have at least one instance of the *fight* theme. Corpus R consists of all the posts that do not contain at least one instance of the *fight* theme.

It can be seen from the first results table below that the density of positive to negative words is lower in corpus F in comparison to corpus R, but the standard deviation of personal pronouns is around four times as great in corpus F in comparison to corpus R.

Table 6-43 density of fight in suicide.net

	posts	tokens	average post length (tokens)	standard deviation of pronouns	ratio positive to negative emotion words
Corpus F fight>0	711 (0.2%)	172,115	242	4.33	.92
Corpus R fight=0	356,228	24,818,964	70	1.28	.99

The table below lists the 50 variables that were found to be significantly different in density between posts in which *fight* is and is not used, ranked according to the proportion of change between corpus F and corpus R. It can be seen that the first-person plural pronouns *we* and *fpp* and the first-person singular pronouns *me*, *my*, and *myself*, are the most increased key variables in Corpus F, with *i* also significantly increased, and with *you* the most decreased pronoun. The emotion terms *Anger*, *Negative*, *Sadness*, *Joy*, *Fear*, *Trust*, and *Positive* are also significantly increased in corpus F.

Table 6-44 most changed linguistic variables between posts that do and do not contain fight

	variable	freq F	freq R		% change	LL	Cohen's d	score
1.	we	3887	2294	+	69.45	156	0.46	1.69
2.	fpp	6257	3804	+	64.51	226	0.47	1.64
3.	me	9993	6262	+	59.59	321	0.57	1.60
4.	VBD	39108	25114	+	55.72	1136	0.51	1.56
5.	my	14862	9930	+	49.67	362	0.58	1.50
6.	myself	1807	1301	+	38.89	30	0.45	1.39
7.	PRP\$	24338	18399	+	32.28	297	0.70	1.32
8.	poss	24780	18886	+	31.21	286	0.68	1.31
9.	compound	10935	15833	-	30.94	291	0.43	0.69
10.	Anger	25053	19623	+	27.67	236	0.74	1.28
11.	<u>you</u>	12707	17558	-	27.63	254	0.45	0.72
12.	intj	2830	3816	-	25.84	48	0.43	0.74
13.	UH	3817	5121	-	25.46	62	0.44	0.75
14.	INTJ	3817	5117	-	25.4	62	0.44	0.75
15.	prt	8343	6705	+	24.43	63	0.64	1.24
16.	RP	8297	6726	+	23.36	58	0.63	1.23
17.	them	2516	2043	+	23.15	17	0.46	1.23

18.	unique_words	513970	665976	-	22.82	6445	0.98	0.77
19.	CD	7722	9622	-	19.75	69	0.41	0.80
20.	NUM	7722	9621	-	19.74	69	0.41	0.80
21.	CCONJ	44238	37359	+	18.41	204	0.76	1.18
22.	Disgust	17552	14844	+	18.24	80	0.68	1.18
23.	conj	44418	38207	+	16.26	164	0.71	1.16
24.	i	50530	43556	+	16.01	181	0.72	1.16
25.	Negative	51227	44277	+	15.7	177	0.76	1.16
26.	Sadness	31857	27627	+	15.31	105	0.73	1.15
27.	PRON	156070	136997	+	13.92	434	0.79	1.14
28.	Joy	21881	19391	+	12.84	52	0.67	1.13
29.	PRP	116713	103684	+	12.57	269	0.78	1.13
30.	Fear	32449	29268	+	10.87	57	0.73	1.11
31.	MD	19958	22385	-	10.85	47	0.64	0.89
32.	it	17024	18996	-	10.38	36	0.7	0.9
33.	ADP	96511	87660	+	10.1	148	0.77	1.1
34.	VBZ	27313	30121	-	9.32	46	0.71	0.91
35.	pobj	73840	67634	+	9.18	94	0.77	1.09
36.	prep	82997	76643	+	8.29	88	0.77	1.08
37.	Trust	30026	27765	+	8.15	31	0.7	1.08
38.	VBG	24989	23145	+	7.97	24	0.75	1.08
39.	neg	20370	18876	+	7.91	20	0.75	1.08
40.	VBP	46422	50152	-	7.44	49	0.84	0.93
41.	Positive	47102	43955	+	7.16	38	0.73	1.07
42.	IN	107260	100635	+	6.58	73	0.77	1.07
43.	RB	88772	83901	+	5.81	47	0.76	1.06
44.	ADV	80743	76541	+	5.49	39	0.75	1.05
45.	monosyllabic words	789832	750189	+	5.28	352	0.81	1.05
46.	VERB	170270	163084	+	4.41	53	0.8	1.04
47.	advmod	78151	74864	+	4.39	24	0.74	1.04
48.	dobj	59722	57262	+	4.3	18	0.78	1.04
49.	NN	127043	131997	-	3.75	32	0.74	0.96
50.	NOUN	156268	161488	-	3.23	29	0.73	0.97

6.8.4 battle

In this analysis the data was split based on the presence or absence of the *battle* metaphor theme. Corpus F consists of all the posts that have at least one instance of the *battle* theme. Corpus R consists of all the posts that do not contain at least one instance of the

battle theme. It can be seen from the first results table below that the density of positive to negative words is slightly lower in corpus F in comparison to corpus R, but the standard deviation of personal pronouns is greater in corpus F in comparison to corpus R.

Table 6-45 density of battle in suicide.net

	posts	tokens	average post length (tokens)	standard deviation of pronouns	ratio positive to negative emotion words
Corpus F battle>0 (0.16%)	559	122,032	218	3.25	0.94
Corpus R battle=0	356,380	24,869,047	70	1.29	0.99

The results table below lists the 47 variables that are significantly different in density between posts in which *battle* is and is not used, ranked according to the proportion of change between corpus F and corpus R. It can be seen that the composite first-person plural pronoun variable *fpp*, which includes the pronouns *we*, *us*, *our*, *ours*, and *ourselves*, is the most increased pronoun variable in Corpus F compared to Corpus R, with *you* the most decreased pronoun variable. The negative emotion variables *Anger*, *Fear*, *Sadness*, *Disgust*, and *Negative*, are significantly increased in corpus F, as are the positive emotion variables *Positive*, *Trust* and *Anticipation*, demonstrating that there is increased emotion talk in conjunction with *battle*. There is no significant change in density of *i* between the two corpora.

Table 6-46 most changed linguistic variables between posts that do and do not contain battle

	variable	freq F	freq R		% change	LL	Cohen's d	score
1.	fpp	6179	3809	+	62.21	150	0.44	1.62
2.	PDT	1934	1332	+	45.18	29	0.46	1.45
3.	their	1852	1395	+	32.74	16	0.42	1.33
4.	Anger	24608	19636	+	25.32	141	0.76	1.25
5.	poss	23469	18905	+	24.15	124	0.70	1.24
6.	PRP\$	22863	18418	+	24.13	121	0.71	1.24
7.	Fear	35696	29259	+	22.00	161	0.79	1.22
8.	WDT	5876	4857	+	20.97	24	0.60	1.21

9.	Sadness	33401	27628	+	20.89	137	0.80	1.21
10.	<u>you</u>	13947	17542	-	20.49	96	0.43	0.8
11.	Disgust	17790	14849	+	19.81	66	0.76	1.2
12.	Negative	52847	44283	+	19.34	189	0.81	1.19
13.	unique_words	543431	665525	-	18.35	2904	0.96	0.82
14.	my	11726	9955	+	17.8	36	0.51	1.18
15.	conj	44488	38219	+	16.4	119	0.76	1.16
16.	compound	13316	15811	-	15.78	51	0.42	0.84
17.	pcomp	9342	8077	+	15.67	23	0.61	1.16
18.	MD	19118	22385	-	14.59	61	0.61	0.85
19.	CCONJ	42669	37381	+	14.15	87	0.8	1.14
20.	it	16381	18995	-	13.76	46	0.66	0.86
21.	amod	45177	39770	+	13.59	85	0.7	1.14
22.	polysyllabic words	76087	67032	+	13.51	142	0.69	1.14
23.	Positive	49856	43948	+	13.44	92	0.74	1.13
24.	relcl	14578	12882	+	13.17	26	0.69	1.13
25.	VBG	26141	23144	+	12.95	45	0.73	1.13
26.	Trust	31131	27764	+	12.13	48	0.73	1.12
27.	attr	15537	13888	+	11.87	23	0.75	1.12
28.	pobj	75546	67638	+	11.69	108	0.78	1.12
29.	VCN	19446	17448	+	11.45	27	0.73	1.11
30.	prep	85314	76644	+	11.31	115	0.78	1.11
31.	long_words	156959	142416	+	10.21	174	0.73	1.1
32.	ADP	96614	87677	+	10.19	107	0.78	1.1
33.	DET	88067	80259	+	9.73	89	0.78	1.1
34.	JJ	68285	62371	+	9.48	66	0.77	1.09
35.	NNS	39301	36103	+	8.86	33	0.68	1.09
36.	ADJ	74472	68560	+	8.62	60	0.78	1.09
37.	DT	80184	74046	+	8.29	60	0.78	1.08
38.	IN	108750	100641	+	8.06	77	0.78	1.08
39.	TO	24395	22630	+	7.8	16	0.77	1.08
40.	Anticipation	30877	28706	+	7.56	19	0.74	1.08
41.	det	67622	63605	+	6.31	30	0.75	1.06
42.	nsubj	113528	120410	-	5.72	49	0.78	0.94
43.	VBP	47651	50138	-	4.96	15	0.82	0.95
44.	NOUN	168808	161416	+	4.58	40	0.72	1.05
45.	VB	59829	62603	-	4.43	15	0.75	0.96
46.	punctuation	125434	120945	+	3.71	20	0.73	1.04
47.	monosyllabic words	767274	750379	+	2.25	46	0.82	1.02

6.8.5 Section summary: metaphor and specific language style

Whereas in cancer.net each metaphoric condition was found to co-occur with language of a style that has been associated with better mental health, for example significantly decreased use of the first-person singular pronoun *i*, and significantly increased use of first-person plural pronouns, this is not the case for suicide.net, in which the work of the different metaphor themes is more diverse. On suicide.net the only indicator of better mental health that is present in every metaphoric condition is the increased standard deviation of pronouns, which however may be partly an artefact of the fact that each metaphoric condition has a larger average word count than does the associated non-metaphoric condition. In the summary table below all notable attributes of language style associated with better mental health are marked bold.

Table 6-47 summary of key changes in each metaphoric condition

metaphor theme	<i>i</i>	first-person plural variables	σ pronouns corpus F	σ pronouns corpus R	corpus F ratio ++ to -- emotion words	corpus R ratio ++ to -- emotion words
journey			2.17	1.29	1.80	0.99
ctb	++		2.40	1.24	1.05	0.99
fight	++	2+	4.33	1.28	0.92	0.99
battle		1+	3.25	1.29	0.94	0.99

It can be seen from the table above that only the *fight* and *battle* metaphor themes are associated with increased use of first-person plural variables, and they are also the conditions in which there is the greatest increase in the range of pronouns used. The first-person singular pronoun *i* is not significantly decreased in any condition and is significantly increased where *ctb* and *fight* are used. The ratio of positive to negative emotion words is only notably increased where the *journey* theme is used. These diverse findings, which contradict the original prediction of the current investigation, that use of metaphor to characterise a concept would co-occur with language of a style that has been

associated with better mental health, are explored in more detail in the diachronic analysis of the next section.

6.9 Diachronic analysis

In this section consideration is given to the change in density of use of the different metaphor themes over time, with use of community metaphors expected to increase in density over time. Subsequently, for each individual metaphor theme, consideration is given to how its most changed key variables relating to state of mind, i.e. personal pronouns and emotion words, change over time, to provide an insight into whether the psychological work of these metaphors changes over time in this community, and what any change may represent. Finally, a comparison is made between different groups of participants over time, to consider whether participation on the forum has an impact on language style as it has been associated with mental health. Each of the 28 months of suicide.net data is considered as a separate data point, with least squares regression calculated for the change over time (by month) for key variables. Where correlation is considered r is interpreted according to the following standard:

Table 6-48 interpretation of r
absolute value

absolute value	relationship between variables
0	none
≥ 0.3	weak
≥ 0.5	moderate
≥ 0.7	strong
1	exact linear

6.9.1 Density of use of metaphor themes over time

In this analysis the proportion of posts containing each metaphor theme per month was considered. The following results table shows the minimum, maximum, mean, and standard deviation of the proportion of posts using each metaphor theme per post, by

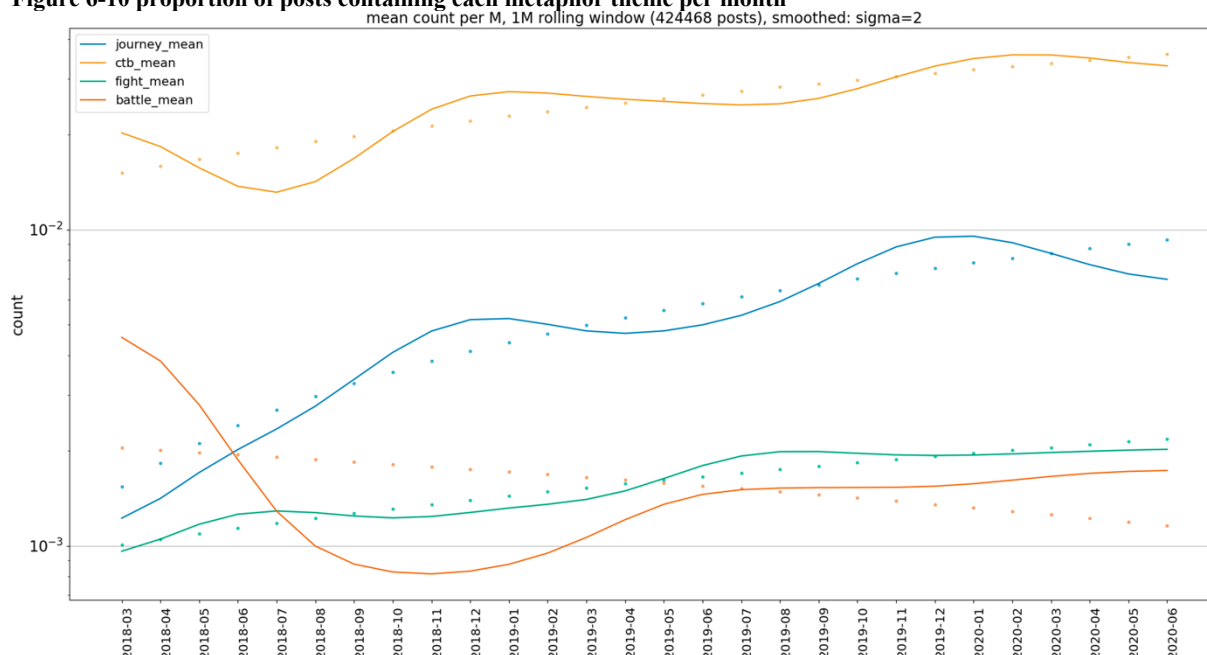
month. For *journey* and *fight* the minimum value is zero, showing that each of these themes was found to be absent from suicide.net in at least one month, with *journey* having zero use in the first two months of this data, and *fight* having zero use in the first month only. The mean values show that *ctb* is the most prevalent theme in the data. There is a strong increase in use of *journey* over time ($r=0.7369$), and a moderate increase in use of *ctb* ($r=0.632$) and *fight* ($r=0.616$) over time, while there is no significant change in prevalence of the *battle* metaphor over time. This supports the interpretation of *journey*, *ctb*, and *fight*, as community metaphors, even though use of *fight* was previously found to be more prevalent in suicide.net.LT20 than in suicide.net.20.

Table 6-49 change in density of metaphor use over time

	min	max	mean	std	r	slope
journey	0	0.0131	0.0054	0.0032	0.7369	0.0002866
ctb	0.0073	0.0413	0.0255	0.0100	0.6320	0.0007693
fight	0	0.0025	0.0016	0.0006	0.6160	4.34e-05
battle	0.0006	0.0096	0.0016	0.0016	-0.1653	-3.28e-05

To show the pattern of change over time, the same data is plotted, below. A least squares regression line (dotted) is plotted for each metaphor theme with the slope described in the table above. The plotted data was smoothed, with sigma set to two, and since the density of the metaphor themes is so different, a log scale was used for the y axis. The *journey* line is blue, *ctb* is yellow, *fight* is green, and *battle* is orange.

Figure 6-10 proportion of posts containing each metaphor theme per month



It can be seen from the plot above that although the decrease in use of *battle* over all 28 months of the data was not found to be significant ($r=-0.1653$), *battle* was the second most prevalent theme at the start of this forum data, but use declined very strongly over the first six months. The strong and steady increase in use of the *journey* theme over time from zero use during the first two months of the forum is also clear.

6.9.2 Correlation of metaphor themes with key variables over time

6.9.2.1 journey metaphor theme

It was shown in a previous analysis that the key variables *your*, *Joy*, and *you* are the most changed in density, each of them increased, in posts containing the *journey* metaphor theme in comparison with posts that do not contain *journey*. The following table shows the minimum, maximum, mean, and standard deviation of r for the correlation of *journey* with each of these most-changed key variables by month, and the change in their correlation with *journey* over time (slope). It can be seen that there is a moderate increase in use of *you* ($r=0.6075$) and *your* ($r=0.6033$) in conjunction with *journey* over time,

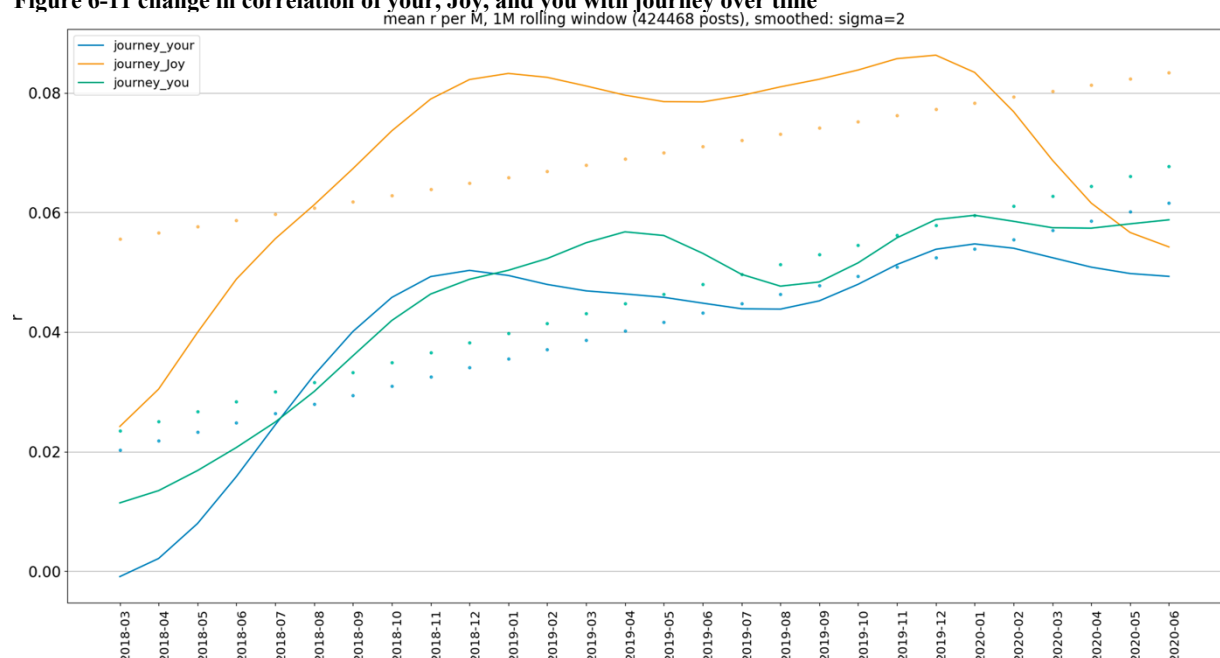
while there is no significant change in the relationship between *Joy* and *journey* over time.

Table 6-50 change in correlation of key variables your, Joy, and you with journey over time

	min	max	mean	std	r	slope
journey_you	0.0	0.0994	0.0456	0.0222	0.6075	0.0016382
journey_your	-0.0141	0.0661	0.0409	0.0209	0.6033	0.0015326
journey_Joy	0.0	0.137	0.0695	0.0329	0.2579	0.0010310

The following plot of the correlation of each of these variables with *journey* over time, where *your* is blue, *Joy* is yellow, and *you* is green, shows for each variable a steady increase over the first eight months, then a levelling off, suggesting that the work of *journey* was established in those first eight months. This suggests again that a particular use of metaphoric *journey* is adopted in suicide.net that can characterise it as a community metaphor, for example the metaphoreme *your journey*. It is also notable that *Joy* comes to be less associated with *journey* after around 23 months.

Figure 6-11 change in correlation of your, Joy, and you with journey over time



6.9.2.2 ctb metaphor theme

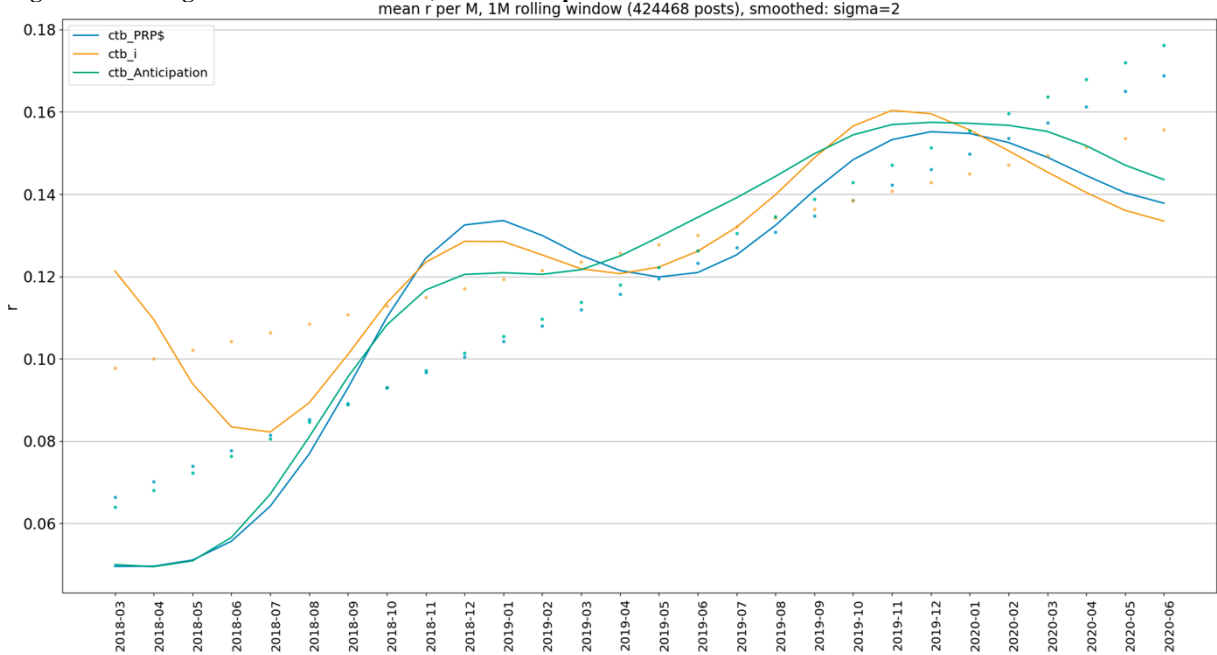
It was shown in a previous analysis that the key variables *PRP\$* (possessive pronouns), *i*, and *Anticipation* are the most changed in density, each of them increased, in posts containing *ctb* in comparison to posts that do not contain *ctb*. The following table shows the minimum, maximum, mean, and standard deviation of *r* for the correlation of *ctb* with each of these most-changed key variables by month, and their change over time. It can be seen that there is a strong increase in use of *Anticipation* ($r=0.773$) and *PRP\$* ($r=0.7263$) in conjunction with *ctb* over time, and a weak increase in the use of *i* in conjunction with *ctb* over time.

Table 6-51 change in use of key variables *i* and anticipation in conjunction with *ctb* over time

	min	max	mean	std	r	slope
ctb_Anticipation	0.0128	0.191	0.1201	0.0443	0.7730	0.0041589
ctb_PRP\$	0.0032	0.1683	0.1176	0.043	0.7263	0.0037925
ctb_i	0.0255	0.2388	0.1268	0.0446	0.3951	0.0021441

The following plot of the correlation of each of these variables with *ctb* over time, where *PRP\$* is blue, *i* is yellow, and *Anticipation* is green, shows a steady increase in the use of *PRP\$* and *Anticipation* in conjunction with *ctb* over time, starting from a near zero relationship, and largely tracking each other, while the relationship between *i* and *ctb* over time is less varied.

Figure 6-12 change in correlation of PRP\$, i and Anticipation with ctb over time



6.9.2.3 fight metaphor theme

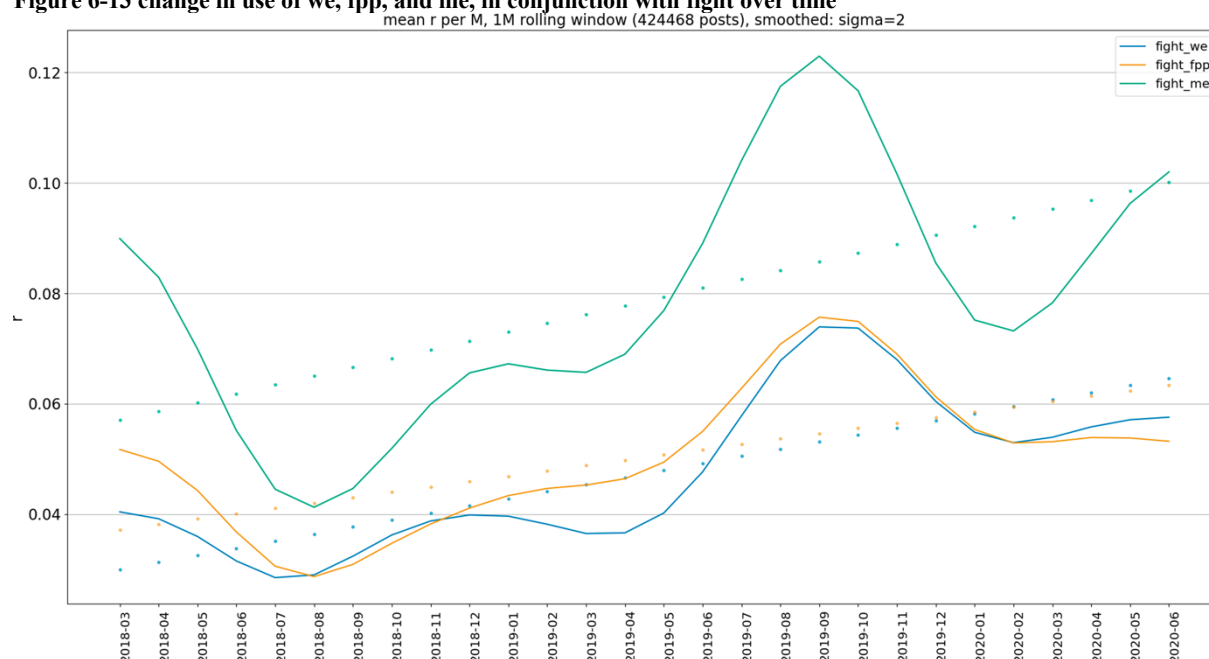
It was shown in a previous analysis that the key linguistic variables *we*, *fpp*, and *me*, are the most changed in density in posts where the *fight* metaphor theme is present compared to posts where it is not present. The following table shows the minimum, maximum, mean, and standard deviation of *r* for the correlation of *fight* with each of these most-changed key variables by month, and their change over time (slope). It can be seen that there is no significant change in the relationship between these three variables and *fight* over time.

Table 6-52 change in use of key variables *we*, *fpp*, and *me* in conjunction with *fight* over time

	min	max	mean	std	r	slope
fight_we	-0.0063	0.1328	0.0473	0.0356	0.2967	0.0012828
fight_fpp	-0.0071	0.1359	0.0503	0.0365	0.2182	0.0009679
fight_me	-0.0087	0.2929	0.0786	0.0678	0.1934	0.0015944

The following plot of the correlation of each of these variables with *fight*, where *we* is blue, *fpp* is yellow, and *me* is green, shows a varied relationship over time, with all three variables largely tracking each other, suggesting that they are typically used together to perform particular work.

Figure 6-13 change in use of we, fpp, and me, in conjunction with fight over time



6.9.2.4 battle metaphor theme

It was shown in a previous analysis that the key variables *fpp*, and *Anger*, are the most changed in density in posts containing the *battle* theme in comparison to posts that do not contain *battle*. The following table shows the minimum, maximum, mean, and standard deviation of *r* for the correlation of *battle* with both of these most-changed key variables by month, and their change over time. It can be seen that there is no significant change in the relationship of *fpp*, and *Anger* with *battle* over time: as with *fight*, the social and psychological work of *battle* on suicide.net remains the same overall, suggesting it is not adapted to do particular work in this community.

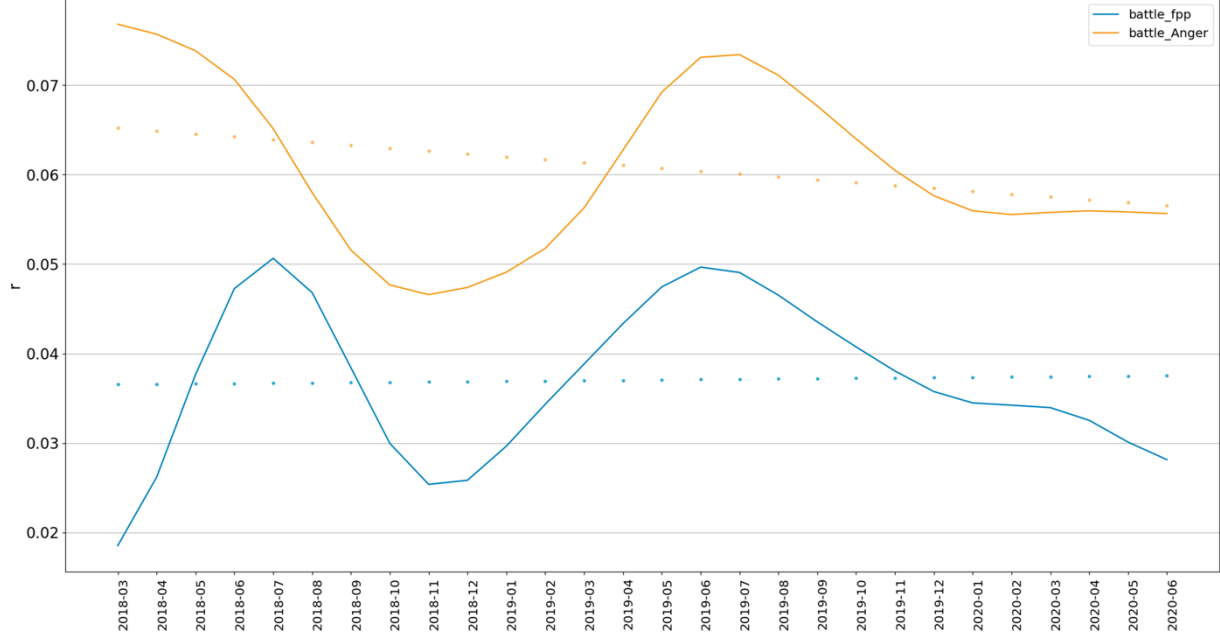
Table 6-53 change in use of key variables fpp, you, and Anger in conjunction with battle over time

	min	max	mean	std	r	slope
battle_fpp	-0.0172	0.1122	0.037	0.0268	0.0112	3.65e-05
battle_Anger	0.0196	0.1609	0.0609	0.0302	-0.0876	-0.0003213

The following plot of the correlation of each of these variables with *battle*, where *fpp* is blue, and *Anger* is yellow, shows that *fpp* increases in conjunction with *battle* over time, starting from a near zero relationship, while the relationship between *Anger* and *battle*

over time is less varied: there is a more consistent correlation between these variables throughout the data.

Figure 6-14 change in use of fpp, and Anger, in conjunction with battle over time
mean r per M, 1M rolling window (424468 posts), smoothed: sigma=2



6.9.3 Change in density of key pronouns for different participant groups over time

To investigate the effect of participation in the suicide.net community on language style, addressed previously in a comparison of keywords between participant groups, in this section consideration is given to the change in density of use of significant terms relating to state of mind for participants in the different stratified participation groups described in the table below. For each group having made more posts, it should be noted that their data also includes the posts their included participants made when they were new to the forum, i.e. each group contains all the data for its component participants, from their first post to their last post.

Table 6-54 comparative participant groups

posts greater than	posts less than	referred to as	participants
0	20	suicide.net.LT20	5,375
19		suicide.net.20	3,213
199		suicide.net.200	477

The results tables for each separate group are listed together below.

Table 6-55 suicide.net.LT20 change in use of key personal pronouns over time

suicide.net .LT20	min	max	mean	std	r	slope
our	0.0002	0.0008	0.0004	0.0001	0.0052	1e-07
we	0.0010	0.0035	0.0016	0.0005	-0.1786	-9.8e-06
i	0.0460	0.0556	0.0517	0.0023	0.0771	2.11e-05
me	0.0049	0.0099	0.0075	0.0011	0.1422	1.91e-05
my	0.0094	0.0126	0.0108	0.0008	0.4225	4.14e-05

Table 6-56 suicide.net.20 change in use of key personal pronouns over time

suicide.net .20	min	max	mean	std	r	slope
our	0.0001	0.0011	0.0004	0.0002	-0.2375	-4.5e-06
we	0.0011	0.0030	0.0021	0.0004	0.4167	2.02e-05
i	0.0375	0.0470	0.0420	0.0017	-0.1094	-2.28e-05
me	0.0051	0.0070	0.0059	0.0004	0.0615	3.2e-06
my	0.0067	0.0094	0.0081	0.0006	0.2860	2.04e-05

Table 6-57 suicide.net.200 change in use of key personal pronouns over time

suicide.net .200	min	max	mean	std	r	slope
our	0.0003	0.0007	0.0004	0.0001	-0.1797	-2.1e-06
we	0.0003	0.0033	0.0022	0.0006	0.4698	3.28e-05
i	0.0330	0.0498	0.0402	0.0034	-0.5102	-0.0002107
me	0.0044	0.0107	0.0058	0.0012	-0.5386	-7.62e-05
my	0.0055	0.0117	0.0074	0.0011	-0.1348	-1.85e-05

It can be seen from the tables above that as participants engage more with the forum their use of the first-person singular pronouns *i* ($r=0.0771$, $r=-0.1094$, $r=-0.5102$), *me* ($r=0.1422$, $r=0.0615$, $r=-0.5386$), and *my* ($r=0.4225$, $r=0.2860$, $r=-0.1348$) decreases, while use of the first-person plural pronoun *we* increases ($r=-0.1786$, $r=0.4167$, $r=0.4698$): a linguistic style that has been associated with better mental health. In contrast to the same analysis for cancer.net, use of the first-person plural pronoun *our* does not increase with increased participation on suicide.net.

6.9.4 Change in density of metaphor themes for participant groups over time

Following on from consideration of the change in density of key pronouns in the previous section, in this section the change in density over time of the four metaphor themes in each participant group is considered, to provide further insight into their identity as community metaphors, with such metaphor expected to be more densely used by more established participants.

Table 6-58 change in density over time for metaphor themes, for participants who have made fewer than 20 posts

suicide.net .LT20	min	max	mean	std	r	slope
journey	0.0	0.0124	0.0047	0.0033	0.3949	0.0001588
ctb	0.0064	0.0515	0.0284	0.0124	0.3983	0.0005988
fight	0.0	0.0056	0.0021	0.0016	0.3692	7.01e-05
battle	0.0	0.0103	0.0016	0.0019	-0.1562	-3.69e-05

Table 6-59 change in density over time for metaphor themes, for participants who have made 20 or more posts

suicide.net .20	min	max	mean	std	r	slope
journey	0.0	0.0131	0.0054	0.0034	0.6925	0.0002832
ctb	0.0056	0.0421	0.0259	0.0101	0.6530	0.0007991
fight	0.0	0.0029	0.0017	0.0007	0.4439	3.84e-05
battle	0.0005	0.01	0.0016	0.0018	-0.2289	-4.89e-05

Table 6-60 change in density over time for metaphor themes, for participants who have made 200 or more posts

suicide.net .200	min	max	mean	std	r	slope
journey	0.0	0.0143	0.0054	0.0035	0.7369	0.0003134
ctb	0.0	0.0415	0.0235	0.0115	0.8112	0.001135
fight	0.0	0.0023	0.0014	0.0007	0.6359	5.27e-05
battle	0.0	0.0035	0.0013	0.0007	0.6233	5.41e-05

It can be seen from the tables above that the density of use of the *journey*, *ctb*, and *fight* metaphor themes increases for each participant group representing increased participation on the forum, with the greatest increases for journey ($r=0.3949$, $r=0.6925$, $r=0.7369$) and *ctb* ($r=0.3983$, $r=0.653$, $r=0.8112$). Density of use of *battle* decreases for the first two groups ($r=-0.1562$, $r=-0.2289$), while it shows a moderate increase over time for the most

established group ($r=0.6233$). This suggests that these comparative groups of participants have a different relationship with the *battle* metaphor. Those who have made > 200 posts may in general have been active on the forum over a longer time period, such that their language style considered together may represent the negotiation and change in the ethos of the forum over time, including use of the *battle* metaphor, which it was shown above was the second most prevalent metaphor theme at the start of the forum before use decreased quite sharply in the first six months of forum activity.

6.9.5 Section summary: diachronic analysis

When all participant groups were considered together the density of use of the *journey*, *ctb*, and *fight* metaphor themes were all found to increase over time on suicide.net, supporting the interpretation that they are community metaphors, while use of the *battle* theme did not change significantly, suggesting that it is not a community metaphor. Use of *you* ($r=0.6075$) and *your* ($r=0.6033$) in conjunction with *journey* increased significantly over time supporting previous insights that *journey* is used to wish other participants well when they *ctb* and suggesting that this particular work of *journey* on this forum develops and becomes more established over time. Use of *i* ($r=0.3951$) and *Anticipation* ($r=0.773$) in conjunction with *ctb* was found to increase significantly over time, suggesting that the work of *ctb* changes, perhaps becoming a way to express personal difficulties and suicidal ideation. In comparison no significant change was found in the key variables associated with *fight* and *battle* over time, suggesting that they are not recontextualised to do particular work in this community.

The diachronic analysis comparing use of key personal pronouns for comparative groups of more and less established participants over time showed that for the most established group of participants (suicide.net.200) there is a weak increase in use of *we* ($r=0.4698$),

and a moderate decrease in use of *i* ($r=-0.5102$), both suggesting an increasing connection with the group. There was found to be a weak increase in use of *my* for more recent participants ($r=0.4255$), which echoes a previous finding that the metaphoric phrase *my journey*, which was found to be used more in line with the more conventional LIFE IS A JOURNEY conceptual metaphor, is used more than three times as often by more recent participants than by more established participants. In contrast to the same analysis for cancer.net, use of the first-person plural pronoun *our* did not change significantly over time for any participant group. The final diachronic analysis showed that use of the *journey*, *ctb*, and *fight* metaphor themes increases with increased participation on the forum, but for the *battle* theme the pattern was found to be more varied.

6.10 Signalled metaphor

In this section metaphor is sought that has not been identified as part of a metaphor theme, on the basis that this may provide insights into the personal lived experience of suicidal ideation, in comparison with dominant metaphor that has evolved over time to perform particular work in the community. It may also provide insights into specific metaphor relating to suicide and suicidal ideation.

6.10.1 Suicide and metaphor

An investigation into the social dynamics of shame and youth suicide specifically considered "the metaphors through which suicide comes to be known" (Fullagar, 2003, p. 291). These were found to include economic metaphors representing suicide as a waste of life and potential. The suicide of young people is described as a rejection of the positioning of life as an object that the entrepreneurial self must maximise in value, a positioning of life that is described as hegemonic in neo-liberal culture.

Suicide in these terms is a matter of becoming waste - the failed or shamed self, the life that did not live up to the expectation of rational autonomous self-management (Fullagar, 2003, p. 292)

The neo-liberal economic metaphors that individualise shame are also described as underpinning the medicalisation of emotional distress through a cultural insistence on the use of psychology experts to solve or fix the dilemmas of modern identity and embodied feeling that underpin suicidal ideation. Metaphors of dirt express the cultural value assigned to young adults marginalised as waste by virtue of their use of public mental health services: if young people do use public mental health services, as society pressures them to do, they are consequently understood as sub-standard, and shameful. The very emphasis on diagnosis and treatment of suicidal ideation, depression, and self-harm as mental health problems, it is argued, may consequently support the process of subjectification in which the self comes to be seen as pathological, and hence shameful (Fullagar, 2003).

Suicide was also found to be metaphorised as an escape to another place, away from the unbearable experience of shame, where shame is described as a response to threatened social bonds. Suicide as escape is a way to replace uncertainty with a sense of purpose, and of achievement, in which death is a place where the unbearable forces of affect no longer apply.

Suicide is the fantasy of laying to rest tumultuous emotions or affective forces generated through a relation to self that is governed by particular expectations about identity (Fullagar, 2003, p. 298)

Use of metaphors of order and chaos similarly show that suicide is understood as a way to overcome the guilt and shame associated with living a disordered life: suicide is a way to

gain control and restore order to their world (Messner and Buckrop, 2000). The metaphoric ideas of escape from disorder, and perfecting the self, are also present in Sylvia Plath's poem *Edge*, written in the days before she killed herself, for example in the lines "The woman is perfected. Her dead Body wears the smile of accomplishment" (Plath, 2021), with the title *Edge* suggesting a transition from one place to another.

In order to identify metaphors for suicidal ideation, and actual suicide, in *suicide.net*, linguistic constructions were sought that are recognised as signalling metaphor. These include similes, copular similes, a sort/kind of, feel as if/though, and the explicit markers 'metaphor', and 'analogy'. Consideration was also given to the co-occurrence of signalled metaphor with the established metaphor themes of *suicide.net*, to investigate whether, as with *cancer.net*, they tend to co-occur – whether use of the metaphor themes engenders use of other metaphor, and thus may facilitate the expression of difficult personal experience and emotion. All of the *suicide.net* data was used in this analysis, and as for *cancer.net*, and as discussed in more detail in chapter three, the spaCy pattern matcher was used to locate signalled metaphor of each type.

6.10.2 Density of signalled metaphor type in different metaphoric conditions

The following table summarises the number of posts in each metaphoric condition, and the number of occurrences of each signalled metaphor pattern found to be used in conjunction with that metaphoric condition. As for the earlier language style comparison, in order to make the non-metaphoric condition more comparable, posts were only included that are at least 10 tokens in length. For the metaphoric conditions *journey*, *fight*, and *battle*, all posts containing those themes were included in this analysis. For the *ctb* and *no-metaphor* conditions, which consist of many more posts, a random sample of

5,000 relevant posts was used. The single copular simile in the *journey* condition was found to be a quote from another post that the participant is replying to.

In comparison to cancer.net, there are insufficient instances to run the chi square test of independence for each signalled metaphor type. However, when all occurrences of signalled metaphor are considered together, the difference in density between metaphoric conditions is significant $\chi^2(0)=802$, $p<0.001$, Cramér's $V=0.47$. This shows that signalled metaphor is more likely to be used where the metaphor themes are present in a post. The order of density of signalled metaphor in metaphoric conditions is *fight*, *battle*, *ctb*, *journey*, *no metaphor*.

Table 6-61 number of posts in each metaphoric condition that contain each signalled metaphor pattern

metaphor theme	posts	as [A] as a [N]	[V+ing] like a [N]	Copular similes	a sort/ kind of	feel as if/ though	metaphor	analogy	total	% posts
fight	711	1	1	4	5	2	2	2	17	2.39
battle	559	0	2	0	0	3	0	1	6	1.07
ctb	5,000	1	4	0	16	4	1	4	30	0.60
journey	2,148	0	1	1	7	1	1	1	12	0.59
no metaphor	5,000	2	2	1	0	2	1	1	9	0.18

6.10.3 Analysis of signalled metaphor

Although Goatly (2011) suggests that similes kill metaphor, the copular similes used in suicide.net were found to be largely creative, and not typically commonly used phrases. And further, although it has been argued that most novel metaphors are creative extensions of existing conceptual mappings (Lakoff and Johnson, 2003) the creative examples discussed below do not just extend conventional metaphor in a new way: they are seemingly personal expressions of the lived experience and worldview of their speaker. In addition, the copular similes and explicit markers on suicide.net, both of which predominantly consist of creative metaphor, were found to naturally form six

groups, which give further insight into the major concerns of the forum. The six metaphor groups are: life; society; mental illness; therapy; suicide; and the suicide.net forum. The signalled metaphor found in suicide.net is now discussed in more detail.

6.10.3.1 Specific warning signs of suicide

Talk of feeling trapped, defeated, or a burden to others, have been shown to be warning signs for suicide (O'Connor, 2021). While metaphors for feeling defeated or a burden are not present in signalled metaphor on suicide.net, the expression of feeling trapped is prevalent, in metaphoric representations of life as a dream, a labyrinth or maze, or a prison:

1. life is like a bad dream you actually can't get out of
2. life is like a cult
3. I feel life is like a labyrinth. the traps are the pleasures in life, and if you fall for them you get stuck in the maze
4. life is like a prison sentence for some but they don't care
5. my life is like an invisible prison where I have treasure a bunch of sharks are constantly trying to capture
6. reality is like an invisible prison
7. i feel as if i'm in jail
8. nearly all the reasons i must ctb is the restrictions people place on me, it really feels as if i'm already in prison
9. what was once a choice has become a sort of sentence
10. existing in this planet as a human, is a sort of prison itself

The feeling of being trapped is also expressed in metaphors of societal structure:

11. the world is like a forced labour camp for most of us
12. reality is like an evil theater where we are forced to be a character suffering in their insane play called life
13. a family is like a miniature enterprise that comes with unavoidable chores and tasks

The expression of feeling trapped is also present in metaphors of mental illness, which specifically reflect the sense of being trapped inside yourself. This is the feeling of being trapped that has been found to be most strongly associated with suicide (O'Connor, 2021):

14. the depressive episodes are like a horrific hole I can't crawl out of
15. it feels as if my soul is trapped inside, screaming to get out
16. i feel as if i'm stuck in this loop this never-ending cycle, that no matter how hard i try it just will never get better
17. mental pain is like an untreatable wound when it gets really bad
18. loneliness is like a void that is eating me alive and i'm drown too deep into the abyss
19. all the feelings are like a never-ending carousel aren't they
20. my mind is like a spinning cd that never stops

To investigate the efficacy of using such signalled metaphors of suicidal ideation to locate other metaphors of suicidal ideation that are not explicitly signalled, a concordance analysis for the term *maze*, discussed above as a metaphor for feeling trapped, found 32 instances in suicide.net, the majority of which do express the feeling of being trapped. In the following example of a metaphorical use of *maze*, suicide is also referred to metaphorically as a means of escape:

it's just being a rat in a maze. at some point, when you're out of options, the only sane thing to do is escape the box

The understanding of suicide as a quick escape from an unbearable situation is also specifically represented in signalled metaphor:

21. i wish our body was like a jump suit and we could just pull the zip down and get out leaving it behind

22. suicide is like a panic button that you can press anytime

23. suicide is like a secret weapon, no rush to use it

In comparison the following metaphor describes suicide.net, rather than suicide, as an escape from feeling trapped:

24. this forum is like a brief escape from it being able to openly talk about what may be considered taboo so you're right, it's one of the few things keeping me ticking over

Metaphors for therapy, in contrast, suggest that therapy is not understood as a solution to feeling trapped, but rather, as was suggested by Fullagar (2003), discussed above, the societal infrastructure around mental concerns is a key part of that feeling of being trapped:

25. counselling is like a mine cart that has gone awry, derailed, and the therapist only seeks to put the mine cart back on track, regardless of the reasoning, the background, and the cost

26. a psychiatrist meddling with the human brain is like a 4-year-old sitting in the cockpit of a spaceship and playing with the controls

27. the therapist is like a robot: you need take medicines, more social relational, stop suicide thoughts etc.

28. mental illness is like a fiat currency: inflation is inevitable, an elder priest determines the value, while it's governed by smoke and mirrors
29. psych wards are like a cattle market, wards being forced to push people out even if the nurses and managers know they are not ready to go
30. antidepressants are like a lotobomy

6.10.3.2 Metaphors for life

Signalled metaphors for life give further insight into the feelings behind suicidal ideation.

In addition to the life metaphors for feeling trapped discussed above, they include metaphors of waste:

1. life is like an abused toilet bowl at a 24-hr restaurant which then overflows, spilling crap everywhere
2. life is like a sewer

metaphors of life as a game, and particularly a gamble, suggesting that the speaker feels a lack of control over their life, which can also be related to feeling trapped:

3. life is like a video game (4 instances)
4. continuing to live life is like a gamble
5. life is like a lottery, if you've got bad genetics well ͡(ツ)͜
6. life is like a casino
7. life is like a game of cards
8. life is like a card game, where we keep getting dealt the shittiest of hands

and metaphors of life as unreal, as a movie, or dream, or story, again suggesting feelings of lack of control, and lack of agency:

9. my life is like a reversed version of fight club
10. my life is like a dull movie
11. my life is like a horror movie
12. even my therapist agrees my life is like a sci-fi movie
13. my life is like a commercial
14. life is like an illusion to me
15. life is like a bad dream you actually can't get out of

6.10.3.3 Metaphors for suicide.net

Metaphors for suicide.net provide insight into the emotional role it plays for participants.

It is a safe haven, a family, a home, a nice meal, a cup of water after walking in the desert, a warm blanket on a cold evening, and an escape. In the one metaphor of negative evaluation it is "like a bad ex that I know I should stay away from":

1. this site is like a safe haven for me.
2. this place is like a safe haven for me to be able to express myself openly without pro-lifers spewing their rhetoric down my throat
3. this site is like a warm blanket on a cold evening, i really cannot say enough about this place
4. this forum is like a brief escape from it being able to openly talk about what may be considered taboo so you're right, its one of the few things keeping me ticking over
5. we'll still be here, and this forum is like a home that you can return to whenever you need/want to
6. this place is like a secret home to me, able to talk about what really bothers me without fear of judgement, detainment, and/or other pro-life spiels

7. this forum is like a family it's really sad that one of the most non-judgmental places i've ever been is a forum for self exiting
8. this community is like a family
9. I do truly feel as if we are a family
10. this place is like a nice meal for me
11. this place is like a cup of water after walking in the desert
12. this site is like a magic luck charm that rubs off on your internet searches
13. i am joining, as long as i can push anybody whom i dislike off the boat right into the atlantic. in a way suicide boat is a nice metaphor for this site
14. we (pro-choicers) are more like apples while (pro-lifers) are oranges, so trying to compare the two and reason with them is like a comparison with apples and oranges
15. this forum is like a bad ex that I know I should stay away from

6.10.4 Section summary: signalled metaphor

There was found to be a significant difference in density of signalled metaphor in the different metaphoric conditions explored, with *fight* and *battle* associated with the most dense use of signalled metaphor, followed by *journey* and *ctb*, and with the no metaphor condition having the lowest density of signalled metaphor. This suggests that using more typical suicide.net metaphor, as represented by these themes, may facilitate other more idiosyncratic expressions of feeling. The order of density of signalled metaphor in conjunction with the metaphor themes also suggests that the themes that have been adapted to do specific work in this community, *journey* and *ctb*, may also constrain discourse in comparison to the less adapted metaphor themes *fight* and *battle*.

Signalled metaphor, and copular similes in particular, located unique creative metaphor in suicide.net that provided insights into the concerns and feelings of participants, as well as making metaphors for suicidal ideation more widely known. In addition, copular similes, and metaphors found through use of the term *metaphor* as an explicit marker, were found to naturally group into the following categories, reflecting the dominant emotional concerns of the forum: life; society; mental illness; therapy; suicide; and the suicide.net forum. The analysis of signalled metaphor on suicide.net provided insights into widespread feelings of being trapped, with societal mental health support framed metaphorically as a part of the trap, while suicide.net itself was almost entirely represented as an escape from those feelings, not through its facilitation of actual suicide, but through the understanding and non-judgmental support it provides to those feeling suicidal, which as discussed above is the primary way that non-professionals can help those who are feeling suicidal.

6.11 Chapter discussion

In this chapter data from a 'pro-choice' suicide forum was used to investigate the triangular relationship between metaphor, community, and state of mind. The keywords of suicide.net in comparison with a general web corpus were found to represent language of a style that has been associated with worse mental health. However, keyword comparison of suicide.net.20 with suicide.net.LT20 found that more established participants have a language style associated with better mental health in comparison to newer suicide.net participants, suggesting that participation on the forum may support improved mental health.

A topic modelling analysis located meaningful interpretative repertoires for suicide.net, including a repertoire associated with the dominant metaphoric term *ctb*, which appears

to relate to planning suicide, including references to dates and occasions. The interpretative repertoire associated with related metaphoric *journey* in comparison shows that it is typically used to wish another participant *peace* and *good luck* on *your journey*, i.e. when they are about to *catch the bus*. Other significant suicide.net interpretative repertoires, which potentially give insights into the concerns of those who are suicidal in general, were found to include:

1. the controversy around suicide as a valid choice
2. existential considerations
3. difficult home life
4. legal considerations around suicide
5. mental health conditions and treatment
6. drugs for suicide
7. issues that prompt suicidal thoughts
8. practical concerns about the act of suicide
9. arrangements for suicide in terms of the place it is done
10. rationale for the experience of suicidal ideation
11. writing suicide notes

Five dominant metaphor themes were identified, with *journey* and *ctb* more dense in suicide.net.20 in comparison to suicide.net.LT20, and *fight*, and *battle* slightly less dense in suicide.net.20, while *struggle* was more notably less dense in suicide.net.20. The *struggle* theme was later shown to be a disparate and potentially less deliberate metaphor, with widely varied uses, and was therefore dropped from further analysis. In another analysis, signalled metaphor, much of which was found to be creative metaphor, was

found to be more prevalent in posts in which the metaphor themes are present, in the order *fight*, *battle*, *ctb*, *journey*, which is exactly the reverse of the relative density of these themes in suicide.net.20 in comparison to suicide.net.LT20. This is summarised in the table below.

Table 6-62 comparative density of metaphor themes by group, and concurrent density of signalled metaphor

metaphor theme	% metaphoric	suicide.net.20 density	suicide.net.LT20 density	suicide.net.20/ suicide.net.LT20	density of signalled metaphor
journey	100	109.15	71.85	1.52	0.59
ctb	100	1,404.69	1,257.72	1.12	0.60
battle	90	22.27	22.69	0.98	1.07
fight	75	36.67	39.33	0.93	2.39

Putting these pieces of information together leads to the insight that for the *ctb* and *journey* metaphor themes that are more truly suicide.net community metaphors in that their use increases over time, and is more adapted over time to do particular work in the community, the community interpretative repertoires of which those metaphors are a part perhaps limit discourse such that there is less space for idiosyncratic metaphoric expressions of feeling, and for the vague language that facilitates the negotiation of meaning. In comparison, the *fight* and *battle* metaphors, which occur slightly less among more established participants, co-occur with signalled metaphor, which includes idiosyncratic creative metaphor, in 2.39% and 1.07% of posts respectively: these more conventionally used metaphors that are less associated with specific community meaning appear to allow more space for, and encouragement of, other figurative discourse. In addition, *fight* and *battle* are the only metaphors to occur alongside increased use of first-person plural variables, an attribute of language style associated with better mental health. At the same time, however, *fight* and *battle* are both associated with increased use of *Anger* words, which as discussed at the start of the current chapter has been linked to suicidal ideation.

That suicide.net use of the community metaphor themes *journey* and *ctb* does not occur alongside language style associated with better mental health is perhaps inevitable given that these metaphors have been adapted in this community specifically to discuss suicide, a negotiation of meaning within the community that also fulfils the meaning criterion for virtual community (Herring, 2004). So, although the *ctb* and *journey* interpretative repertoires appear to constrain discourse, at the same time adoption of the community recontextualisation and extension of the hegemonic *journey* metaphor as the *ctb* theme both provides a vehicle for discussion of suicide, and simultaneously creates a sense of belonging. This finding must also be considered in the very specific context of suicide.net in which there is the inherent potentially paradoxical task of demonstrating genuine suicidal ideation while at the same time becoming a member of the suicide.net community.

The *journey* metaphor as it is used in suicide.net appears to recontextualise the hegemonic LIFE IS A JOURNEY conceptual metaphor. In the context of suicide.net the *journey* metaphor loses the sense of being a long-term endeavour over a lifespan and becomes an imminent short-term event in which someone ends their life. The subversion of the hegemonic LIFE IS A JOURNEY metaphor into the *your journey* metaphoreme is reflected in the distinctive language style surrounding use of *journey*, which both types of supervised learning algorithm predicted the most accurately, with high weighted average F1 scores of 0.81 and 0.86. In comparison it was shown that for the more recent participants the phrase *my journey* is more prevalent, while a concordance analysis shows that use of *my journey* is still consistent with the more conventional lifespan orientation of LIFE IS A JOURNEY: it has not yet been recontextualised.

A similar recontextualisation occurs with the *rollercoaster* metaphor which was found in the cancer.net analysis to have a high word vector similarity to *journey* due to its related metaphoric use in general discourse. Since *rollercoaster* is present to a much lesser extent in suicide.net - there are 127 instances, many of which are literal - it has not been a focus of the suicide.net investigation. Rather than being used in the typical metaphorical manner to represent the short-term ups and downs of emotions and experience, on suicide.net the *rollercoaster* metaphor may typically be recontextualised to express only downward motion, and the sense of being trapped, for example: "i think my rollercoaster went under water and never came back up"; and "i'm tired of that down spiral roller coaster". The unexpected subversion of the anticipated conventional use of *rollercoaster* is stark and expressive.

In contrast to cancer.net, where participants share the common experience of living with cancer, there is no obvious overriding common attribute of participants on suicide.net, except perhaps a lack of significant other community. Community on suicide.net, which has been shown to meet the criteria of a virtual community (Herring, 2004), comes from a more basic existential aspect of being human, for example as described by Lingis (1994) in *The Community of Those Who Have Nothing in Common*:

In the midst of the work of the rational community, there forms the community of those who have nothingness, death, their morality, in common (Lingis, 1994, p. 13)

The Interpersonal Theory of Suicide (Joiner, 2007) states that suicidal desire emerges from three factors: (i) the feeling of being a burden to others; (ii) feeling that you don't belong; and (iii) the learned ability to hurt yourself. Change in pronoun use by participants over time suggests that participation on suicide.net may in general help to alleviate the sense of not belonging. Specifically, as participation increases, use of the

first-person plural pronoun *we* is increased, and use of first-person singular pronouns is reduced. In addition, expressive idiosyncratic metaphor frames suicide.net as a safe haven, a family, a home, a nice meal, a warm blanket on a cold evening, and an escape. While idiosyncratic metaphor suggests that suicide.net is experienced as helpful in dealing with the feelings of being trapped that have been found to be most representative of suicidal ideation (O'Connor, 2021), metaphor also provides the insight that conventional societal structure around mental health may contribute to feelings of being trapped. As well as supporting the findings of Fullagar (2003) discussed above that societal mental health services may contribute to feelings of shame and suicidal ideation, this also aligns with research showing that psychiatric hospitalisation in general increases risk of suicide for many years after discharge (Chung *et al.*, 2017), with adverse experiences in hospital potentially the cause of some post-discharge suicides (Chung, Ryan and Large, 2016), and with patients who perceived themselves as having been coerced into hospitalisation found to be significantly more at risk of killing themselves both during treatment and after their discharge, with confounding factors taken into account (Jordan and McNiel, 2020).

A report on mental health and human rights by the United Nations High Commissioner for Human Rights advises a comprehensive approach to human rights for those with psychosocial disabilities, including protection of their autonomy, agency, and dignity, as well as requiring policy shifts that recognise exclusion and marginalisation as the causes and consequences of poor mental health (United Nations High Commissioner for Human Rights, 2017). Linguistic investigation in the current chapter suggests that suicide.net may help support all of these rights, while the inability of hegemonic society to tolerate suicide.net (Closure of suicide forums - Early Day Motions - UK Parliament, 2020)

correspondingly tears them down. For example, from a Reddit discussion following the sudden closure of R/sanctionedsuicide:

This will only end in more suicide. They know it, but hey who cares about tearing apart a community of people that otherwise don't have one as long as their asses are covered

6.12 Next chapter

In the concluding chapter of this investigation, data about metaphor use and language style in cancer.net and suicide.net is considered together, and conclusions are drawn about the social and psychological work of metaphor, and the relationship between metaphor, community, and state of mind.

7 DISCUSSION AND CONCLUSION

7.1 Discussion

The aim of this thesis was identified in chapter one as the investigation of the triangular relationship between metaphor use, community, and state of mind, such that insights would be provided into (i) the effect of metaphor use in terms of state of mind; (ii) the role of metaphor in the characterisation of a community; and (iii) methods for considering linguistic metaphor in naturally occurring discourse in terms of its psychological effect. The investigation compared two large corpora of naturally occurring, computer-mediated discourse, each consisting of all posts from an active and long-term internet forum established to address a particular life-threatening experience. Forum one, referred to as cancer.net, is focused on the experience of living with cancer, and forum two, referred to as suicide.net, is a 'pro-choice' suicide forum, where 'pro-choice' means that suicide is accepted as a valid choice.

On the basis that more instinctive metaphor is not so available for the characterisation of a community, the focus was on potentially more deliberate metaphor themes, and metaphoremes, as they are instantiated as nouns. For cancer.net the identified metaphor themes were journey, rollercoaster, fight, and battle, with *cancer journey* identified as a metaphoreme which was shown to be dominant in conjunction with multiple metaphor themes, not just the journey theme, such that it may support an ideological dilemma. For suicide.net the identified themes were journey, ctb, fight, and battle, with *your journey* identified as a metaphoreme used to convey support for the suicide of another participant, an antithetical recontextualisation of the LIFE IS A JOURNEY conceptual metaphor.

This investigation has shown that for the cancer.net community use of the journey metaphor theme is (i) the most increased over time in terms of the proportion of posts it is

present in; and (ii) increasingly more densely used in journey posts over time; and in addition participant use of the journey theme is the most increased with increased participation on the forum. In comparison use of the fight and battle themes is significantly decreased over time on cancer.net as a whole, and use of the fight and battle themes decreases over time for all groups of participants compared based on number of posts made, indicating that these themes are inhibited on cancer.net in a way that is communicated to participants from the very start of their participation. For suicide.net also use of the journey metaphor theme is the most increased over time, a pattern that has strong significance ($r=0.7369$), while there is a moderate increase in use of the ctb and fight themes over time. The journey theme and the community specific ctb (catch the bus) theme, which the community has forked from journey to do specific, related work, are the two metaphor themes for which participant use is most increased with increased participation on the forum, based on comparative participant groups. For suicide.net use of the fight metaphor theme also increases over time and with increased participation on the forum, but for the battle theme there is an overall decline in use, although this is not significant, and the pattern for the between group comparison is more mixed.

The pursuit of meaning in life is considered within psychology to be a fundamental human motivation (Frankl, 2004), such that perceived meaning in life predicts good mental health and social functioning, while a sense that life is meaningless has been associated with compromised functioning, and pathological states, including depression, and suicidal ideation (Harlow, Newcomb and Bentler, 1986). Situational reminders of death, which are prevalent in the discourse of both the fora investigated in the current study, have been found to decrease meaning in life, and this is particularly the case for those with low self esteem (Routledge et al., 2010). Landau (2017) identifies the general dominance of the LIFE IS A JOURNEY and LIFE IS A STORY conceptual metaphors as

relating to self-relevant motives of self-esteem, self-continuity, and self-growth, which have been psychologically important cross-culturally and historically, in that the sense of autobiographical continuity they entail defends against fear of death (Becker, 1997). It is perhaps unsurprising then that for both communities it is the LIFE IS A JOURNEY conceptual metaphor that has been adopted and adapted to address a particular existential threat (Yalom, 1980) which entails the potential end of life in the short term, which for cancer.net participants is largely beyond their control, while for suicide.net participants in comparison their death in the short term is ostensibly within their control.

On cancer.net the journey metaphor theme was found to frame the experience of living with cancer in the timeframe of a life, such that it may become more consistent with the longer-term outlook that is likely to have underpinned the conceptualisation of life before the experience of living with cancer began. In comparison the battle metaphor on cancer.net was found to be predominantly used in association with specific periods of time, including day, week, month, half year, and year, but not life, and this was the case in 70% of uses of battle for more recent participants. This may be related to the conceptualisation of time as "a river or a conveyor belt on which events are moving from the future to the past" (moving time) (Boroditsky, 2000, p. 5), in comparison to a conceptualisation of time in which the "'ego' or the observer's context progresses along the time-line toward the future" (moving ego) (Boroditsky, 2000, p. 5). Cancer is typically a taboo subject, with use of cancer as a metaphor, which may be said to represent the hegemonic folk psychological conceptualisation of cancer, always invoking "hard-to-control processes described as growing, spreading, or eating away" (Potts and Semino, 2019, p. 93). When the experience of cancer begins it may feel to the participant as though cancer has intruded on their life, since they surely do not seek, or typically anticipate, cancer, such that they may respond in terms of a more immediate defensive

cancer battle of a day, week, month, half-year, and year. The *cancer journey* metaphoreme in comparison may be said to reframe the experience of cancer into the longer term, such that cancer is typically not the sudden, catastrophic cliff edge, that it might appear to be when first encountered, but is consistent with the difficulties that are inherent in any life journey. The *cancer journey* framing invokes the moving ego perspective inherent in the hegemonic LIFE IS A JOURNEY metaphor, such that it may support an increased sense of agency and coherence in conjunction with the experience of cancer.

For the suicide.net community there is no notable use of particular time references in conjunction with its metaphor themes, but there is a notable change in time perspective related to metaphor. On suicide.net, which is based on consideration of the deliberate cessation of life in the near future, the LIFE IS A JOURNEY conceptual metaphor has been transformed and extended to represent a short-term perspective based on the mundane and accessible activity of catching the bus, while the rest of the community wishes them well on their journey. This is consistent with the finding that the further the temporal distance to an event, the more likely it is to be represented in terms of a few abstract features that convey the perceived essence of the events, for example the wider cultural instantiation of the LIFE IS A JOURNEY conceptual metaphor, rather than in terms of more concrete and incidental details of the events, in the case of suicide.net the reduction of LIFE IS A JOURNEY to, for example, buying tickets for the bus, having a bus pass, waiting for the bus, and arrival and departure times for the bus. The difference in the level of construal of events that this represents, i.e. abstract to concrete, entails a corresponding psychological distance to events, i.e. more distant to more close, which has been found to have a broad range of behavioural consequences (Trope and Liberman, 2003; Jia and Smith, 2013). This effect may be compounded in that between person

differences in subjective health, loneliness, and life satisfaction at a fixed point in time, as well as within person changes across time in subjective health and life satisfaction, have been found to affect future time perspective (Korff and Biemann, 2020).

It has been shown that on cancer.net the journey metaphor theme co-occurs with language of a style that has been associated with better mental health. However, for suicide.net the journey metaphor, and the specific community extension of journey into the ctb metaphor, have been reconceptualised to support suicide discourse, and consequently do not co-occur with a language style that has been associated with better mental health. The suicide.net community is based on a stance towards life that is orthogonal to that of the wider community. Within the suicide.net community, then, LIFE IS A JOURNEY is reconceptualised to represent life as the opposite of precious, such that the ending of life is a preference.

we're institutionalized to life. we don't know, we don't want to know, and we will not believe what's on the other side (from suicide.net)

However, although on suicide.net the recontextualised journey metaphor is not used in conjunction with language of a style that has been associated with better mental health, in their communal act of recontextualisation of the hegemonic LIFE IS A JOURNEY metaphor, the identity of suicide.net as a community is strengthened, and paradoxically this may simultaneously support a stance towards life that is more consistent with the hegemonic view. This also suggests that the recontextualised meaning and associated work of the journey metaphor on suicide.net may still be understood in relation to, and simultaneously with, the hegemonic understanding of LIFE IS A JOURNEY. It was surmised that the suicide.net journey metaphor theme, and its related idiosyncratic community specific ctb theme, perform an essential function in supporting the

conceptualisation and discussion of suicide, such that, as was shown in the investigation, the more a participant engages with the suicide.net community, the more their language is of a style that has been associated with better mental health. This is consistent with the insight discussed previously that the best way that non-professionals may help those who are feeling suicidal is to listen without judgment to their concerns, and without trying to 'fix' their personal existential crisis (O'Connor, 2021). It is also consistent with the finding that for participants with low perceived levels of coherence in their life, a journey framing increased their sense of purpose in comparison to a no-metaphor condition, and this effect was increased where participants generated a life metaphor of their own in comparison to a supplied journey framing, which included idiosyncratic metaphor without a temporal dimension (Baldwin, Landau and Swanson, 2018).

The use of the different metaphor themes in combination, either in the same posts, or in the same interpretative repertoires if not in the same posts, was related to the analytic concept of ideological dilemmas, which as discussed in chapter one may be understood as enabling meaningful thinking about the self and the world. In cancer.net the mixing of the journey metaphor theme with other metaphor themes was found to be specifically associated with use of the first-person plural pronoun *our*, suggesting that subject positions in relation to the experience of cancer are worked out in interaction with the group. For example, in saying "it's not about winning or losing this journey of ours" (Table 3-3) there is an (authoritative) implication that (i) a fight or battle framing, represented by the phrase *winning or losing*, which in terms of the diverse experience of cancer is necessarily related to the individual rather than the whole group, is dispreferred; and (ii) we, this community, are all still on our (life) journey. In addition, it was found that the first-person singular pronoun *I* was significantly decreased in conjunction with the fight and battle themes considered across all uses, but *I* was not significantly

decreased when fight and battle were each considered in isolation from their use with other metaphor themes. This further supports the insight that on cancer.net metaphor mixing is related to the negotiation of subject positions in relation to the experience of cancer, with fight and battle invoking a framing that relates more to the individual, while journey relates to the shared experience and ethos of the community.

Gramsci (2007) states that in any particular situation a particular rhetoric, or rhetorics, may be preferred, or hegemonic, such that they assume the status of facts. In this way, people become the products of discourse as well as producers of it (Edley, 2001). On cancer.net the journey metaphor framing comes to dominate emotion talk, and at the same time prioritises positive over negative emotion talk. It has been shown that positive affect may predispose people to feel that life is meaningful, and may also increase sensitivity to the meaning relevance of a situation (King *et al.*, 2006), and that this is more the case as future time becomes limited (Hicks JA *et al.*, 2012). However, at the same time the journey metaphor theme on cancer.net may limit the full psychological openness of participants to their condition. For example the MCEL study found that dominance of a journey metaphor may reflect and reinforce a sense of powerlessness, loneliness, and the inability to express negative emotions such as frustration, dejection, and lack of acceptance (Semino *et al.*, 2017). Another study found that where healthcare professionals encourage cancer patients to 'fight', and have a positive attitude, they effectively promote and collude with emotional suppression of this difficult experience (Byrne *et al.*, 2002). On suicide.net similarly the journey and related ctb metaphors are used to facilitate suicide discourse such that, as discussed above, they may promote a shorter-term perspective on life. And while it was shown that for cancer.net the identified metaphor themes are more likely to occur with other, signalled, metaphor, including creative metaphor that may facilitate the expression of more personal, subjective

experience, and vague language that may support the negotiation and evolution of meaning in conjunction with metaphor, for the suicide.net forum in comparison the more a metaphor theme has been adapted for specific work in the community (journey and ctb) the less likely it is to occur with signalled metaphor, such that on suicide.net these aspects of discourse may be constrained by those dominant metaphor themes.

The current investigation, then, supports the view of metaphor as a powerful rhetorical device which may strongly characterise a particular discourse community such that the actions of its members may be corralled and constrained in both positive and negative ways. And further, the recontextualisation of dominant metaphor from the wider culture in a particular community may compound and amplify the dominant, potentially orthogonal, ethos within that community, for example in an echo chamber effect, in which beliefs are reinforced by their repetition and amplification within a closed system, which has been much discussed in conjunction with the internet (Garrett, 2009; Kelly, 2019; O'Hara & Stevens, 2015; Schaub & Morisi, 2020; Vaccari, 2013). However it has been found that individuals are fairly consistent in their tendency towards literal or metaphoric thought, and alongside increased susceptibility to metaphoric transfer effects, a tendency towards more metaphoric thought has been associated with higher levels of emotional understanding (Fetterman et al., 2016), which further supports the importance of metaphor in these communities which exist to support specific emotionally challenging experiences. It was surmised above that conceptual metaphor may support multiple understandings, such that metaphor recontextualised to do particular work in a community may be understood in relation to, and simultaneously with, the hegemonic meaning from which it is forked. For example Tay (2011a) uses a case study of the conceptual metaphor THERAPY IS A JOURNEY across different levels of therapeutic discourse to show that the inherent stability afforded by conceptual metaphors supports a

sense of theoretical continuity, while simultaneously allowing variations to arise out of situated discourse objectives. It has been argued that such contextual re-interpretations of the conflictual meanings inherent in complex metaphor are "the most typical and revealing instances of living metaphor" (Prandi, 2012, p. 148).

In summary, the ethos on each forum both creates the participants and is created by them, such that each shapes the other over time, and for each forum the recontextualisation of the pervasive and powerful LIFE IS A JOURNEY conceptual metaphor is the primary vehicle for this change. The primary prediction of this thesis was that where metaphor is used to characterise a concept the surrounding language will be of a style that has been associated with better mental health, and this is true for the identified metaphor themes of cancer.net. For suicide.net in comparison the LIFE IS A JOURNEY metaphor has been adapted by the community for a community-specific purpose that entails its association with a language style that has been associated with worse mental health. Paradoxically that in itself seems to be protective in general: because they use the characteristic community metaphor to represent their shared concerns, participants are part of the group, they share the experience of the group, and this in turn legitimises their own experience, and helps to defend against their sense of alienation from wider society. For both cancer.net and suicide.net, being part of the specific community that resonates with their own pressing concerns and experience helps them to actualize their identity, ideology, and agency in society (Morgan, 2014), such that use of language of a style associated with better mental health, which is based on a subject position of membership entitlement, increases with increased participation in each community. And increased participation in and membership of the community is itself characterised via recontextualisation of the dominant LIFE IS A JOURNEY conceptual metaphor.

7.2 Conclusion

7.2.1 Contributions to new knowledge

To address the original aims of the current study, then, it has been shown that use of the identified dominant metaphor themes in each community co-occurs with specific language styles that have been associated with state of mind, and that this work of metaphor evolves over time as a consensus which becomes normative within the group for a period, while the flexibility of metaphor still leaves that work open to further evolution. The adaptation and promotion of particular metaphor themes over time to do particular work in each forum also contributes to the characterisation of it as a particular community.

In terms of metaphor theory, the identification of the different psychological work of metaphor in these contexts from the identification of linguistic metaphor supports the conceptual metaphor theory approach that linguistic metaphor is underpinned by cognitive action. This issue is represented in the primary research question of the thesis, whether where metaphor is used to characterise a concept, the surrounding language is of a style that has been found to be associated with better mental health. This was found to be the case for cancer.net, but specifically was not the case for suicide.net. However, for both corpora investigated use of the identified metaphor themes to characterise a concept was surrounded by distinctive language of a style that can be connected to mental health based on established research findings (section 1.4). In the current investigation this issue was considered in relation to potentially deliberate metaphor, which was related to deliberate metaphor theory (sections 1.3.3, 3.2.3). This analytic decision was made on the basis that it is unlikely that more pervasive instinctive metaphor, i.e. *metaphors we live by* (section 1.3.1.3), such as the conduit metaphor (Reddy, 1979), may be corralled to have

particular work hitched to it within a community. The recontextualisation of the identified metaphor themes for community purposes, and their use together, also supports interpretation of them as potentially deliberate, such that there is communicative intent behind their use (sections 1.5.3.1, 3.2.5). The data of the current study consequently can not be generalised to the *metaphors we live by* on which conceptual metaphor theory (Lakoff & Johnson, 2003) is focused.

The current study also supports the understanding of metaphor as dynamic (section 1.3.4), in that density of use of the metaphor themes, and their relationship with key variables related to language style, have been shown to be different in the different fora, and to change over time, including for comparative participant groups. The focus in the current study on word counting, and metaphor instantiated as nouns, which has been shown in the supervised learning investigations in sections 4.2 and 6.7 to be highly effective in supporting the aims of the current thesis, may be said to be inconsistent with the discourse dynamics approach (sections 1.3.4.1, 1.3.5). The overarching discursive psychology approach, however, is consistent with discourse dynamics, and in section 1.5.2.1 the discourse dynamics concept of systematic metaphor is specifically related to the discursive psychology analytic concept of interpretative repertoires. Considering ideological dilemmas as a specific site of work and evolution of meaning (section 1.5.3.1) is also consistent with the talking and thinking analytic approach to metaphor represented in the discourse dynamic framework concepts of metaphor shifting, and systematic metaphor (Cameron et al., 2009; Cameron, 2007). Concepts from the discourse dynamics approach are also considered in the exploration of signalled metaphor in sections 3.5 and 6.10. Vague language such as *a sort of*, and *a kind of*, is specifically excluded from the MIPVU metaphor identification procedure which, as discussed in section 3.2.5, is more suited to the identification of metaphor as it is operationalised in the current study.

However, the view that such language has a strong relationship to metaphor, including potentially deliberate metaphor (Cameron & Deignan, 2003, 2006), particularly in terms of how the vagueness of metaphor allows it to be adapted for different purposes over time (Cameron, 2007), is supported by the current study.

The mixing of metaphor also is related to deliberateness in the current investigation. As discussed in relation to ideological dilemmas (section 1.5.3.1) the mixing of metaphor has been argued to be an instantiation of potentially more deliberate metaphor (Steen, 2016). The perception that metaphors are being mixed, it is argued, depends on the interpretation or experience of those metaphors as deliberate: "all other cases of conceptual clashes between adjacent metaphors do not get recognized as mixed metaphor because they are not used deliberately as metaphors" (Steen, 2016, p. 114). From a dynamic perspective this may be related to activation levels of metaphor, where "the mixing of metaphors changes the semantic salience structure, creating different version and degrees of activated metaphoric meaning" (Müller, 2016, p. 32). In comparison to the discourse dynamics approach (Cameron et al., 2009), which has a focus on potential metaphoricality of linguistic metaphors for participants in a conversation, Müller (2016) relates the mixing of metaphor to display by participants of the presence of metaphoric meaning in their communicative intent. In terms of the career of metaphor theory (Bowdle & Gentner, 2005), the mixing of metaphor perhaps reinstates cross-domain mapping (comparison) when in metaphor in comparison to simile there is typically within-domain mapping (categorisation), which may be more conventional, and therefore potentially less deliberate. In terms of health discourse specifically, Charteris-Black (2016) similarly finds that the mixing of metaphor supports the communication of pain, specifically occurring where the intent is to emphasise the intensity of pain, and to characterize it as out of control. Where aspects of pain are referred to as controllable, in comparison,

metaphor use is more semantically convergent, such as repeated and extended metaphors based on the same source domain. The potential deliberateness of metaphor mixing was supported in the limited investigation into this area in the current study in relation to the cancer.net corpus, in which it was found that the *journey* metaphor theme is significantly more likely to be used in conjunction with the first-person plural pronoun *our* when it is used in conjunction with other identified metaphor themes, than when it is used in isolation. This suggests a deliberate communicative intent, which may be related to subject positions within this community. For example, as discussed above, in saying "it's not about winning or losing this journey of ours" (Table 3-3) there is an (authoritative) implication that (i) a fight or battle framing, represented by the phrase *winning or losing*, which in terms of the diverse experience of cancer is necessarily related to the individual rather than the whole group, is dispreferred; and (ii) we, this community, are all still on our (life) journey. Metaphor mixing may also be related to blending theory (section 1.3.2), in which multiple conceptual structures are combined, which may be essential to understanding complex mixed metaphor. While conceptual metaphor theory states that mixed metaphor is impermissible, such that the findings of the current study which demonstrate that the mixing of metaphor has a strong effect on language style as it relates to state of mind might undermine that conceptual metaphor position, it is argued in section 1.3.2 that in addressing different aspects of the conceptualisation of metaphor, blending theory and CMT are largely complementary, where blending theory may be essential to the explanation of how complex mixed and extended metaphor may be understood. Similarly, aspects of all metaphor theory discussed in chapter one have been supported in the findings of the current study.

In terms of metaphor identification, consideration of language style in conjunction with metaphor use may be a productive method of finding metaphor in large text corpora; for

example a supervised model accurately predicted the presence of metaphor in posts labelled as not containing metaphor, where the posts did contain evocative, potentially deliberate, new metaphor, which was just not part of the identified dominant metaphor themes (section 4.3.4.2.1). This points to the social and psychological work that such more deliberate or active metaphor supports in general, in that it is surrounded by a particular distinctive language style. The signalled metaphor analysis in sections 3.5 and 6.10 was also productive in finding additional metaphor. This analysis created valuable insights into the experience of participants in relation to the health conditions of focus of each forum investigated. It also created some insights into metaphor theory. The signalled metaphor analysis demonstrated that similes, which Goatly (2011) suggests kill metaphor, were unique in 87% of uses (section 3.5.5), and were predominantly creative, and not typically used phrases, which is consistent with the career of metaphor theory discussed in conjunction with deliberate metaphor theory in section 1.3.3. The signalled metaphor analysis also showed, as discussed above, that the vague language *a sort of* and *a kind of*, which is specifically disallowed in the MIPVU metaphor identification procedure, has a strong relationship with the identified metaphor themes, and this supports further investigation into the role of vague language in conjunction with metaphor, and the adaptation of metaphor over time.

While the current study has supported these insights into metaphor theory, the methods of the study are relatively narrow, focusing, as discussed above, on prevalent potentially deliberate metaphor from interpretative repertoires, as they are instantiated as nouns, and taking a predominantly quantitative approach. It has already been noted that although the discourse dynamics approach is not consistent with this method of the current study, concepts from the discourse dynamics approach have been productive in providing analytic insights. As was discussed in chapter one, the metaphor theory addressed in the

current study builds on conceptual metaphor theory, such that each theory is related in fundamental ways. The data of the current study, which has been linked to insights from each of the theories, supports this understanding. So, while the method and analytic focus of the current study does not address the whole range of metaphor, and does not address theory outside of the conceptual metaphor tradition, it does provide some productive insights, and it does support the understanding of metaphor as a very powerful social and psychological mechanism.

The final contribution of the current investigation is the exploration of methods from the areas of natural language processing and data science to use alongside more traditional corpus linguistic techniques. These have proved productive in the following ways:

1. finding metaphor, including signalled metaphor, and identifying metaphoremes;
2. locating the characteristic language and concerns, the interpretative repertoires and ideological dilemmas, of each community;
3. validating the concept that individual metaphor themes, and all metaphor themes together, are surrounded by a specific language style representing the social and psychological work that they do, such that their presence in a post can be predicted with high accuracy by a supervised learning model;
4. processing the text in general for use in the other investigations of the study.

In terms of wider issues, the current investigation provides insights into the ongoing debate around the internet as a site for health support and has particular relevance for the current active debate about the existence of suicide.net (*Closure of Suicide Forums - Early Day Motions - UK Parliament*, 2020). As discussed in chapter six, the United Nations High Commissioner for Human Rights advises a comprehensive approach to human rights for those with psychosocial disabilities, including protection of their autonomy, agency, and dignity, as well as requiring policy shifts that recognise exclusion

and marginalisation as the causes and consequences of poor mental health (United Nations High Commissioner for Human Rights, 2017). The current investigation suggests that suicide.net may help support all of these rights, and this knowledge is new, and of vital importance.

7.2.2 Limitations and future work

In the topic modelling analysis of the current study (sections 2.8, 6.4.1), topics are considered via the discursive psychology analytic concept of interpretative repertoires (section 1.5.2), and the analytic focus is on interpretative repertoires in which potentially metaphoric terms are present. The process of identification of potentially metaphoric terms at this stage should be more rigorous, with each extracted term in each topic considered for actual metaphoricity in its context. This issue is addressed to some degree by the use of word vectors to identify additional potentially metaphoric terms semantically related to potentially metaphoric terms identified by the researcher from interpretative repertoires (sections 2.9, 6.4.2). The use of word vectors with the suicide.net data, for example, identified the potentially metaphoric term *struggle*, which was present in two interpretative repertoires but not identified as potentially metaphoric by the researcher. While consideration of *struggle* as a metaphor was later discontinued on the basis that it does not appear to invoke a particular source domain strongly (section 6.6.7), this issue remains a weakness of the current method which could potentially miss an important metaphoric term.

In the diachronic analyses of the current study (chapter 5, section 6.9) linear methods are used to consider change over time of metaphor use and key linguistic variables relating to the research questions of this thesis. Tay (2019) notes that the issue of autocorrelation, which relates to the presence of internal relationships within temporally ordered

observations, is typically overlooked in corpus linguistic research, in which independence between time periods is assumed, and possible interdependence between temporally ordered observations is not considered. Tay suggests time series analysis as a complementary methodological approach that can address such limitations, which may provide new structural insights into discourse behaviour over time, including potentially the forecasting of future values. Time series analysis is not undertaken in the current analysis, which focuses on trends across the whole data. While plots of the diachronically focused analysis carried out in the current investigation help to address this issue, allowing the visual inspection which has a fundamental role in time series analysis (Koplenig, 2017), specific time series analysis would complement this approach and should be undertaken in future analysis.

In the current study the mixing of metaphors, which it has been argued is inherently more deliberate and potentially related to communicative intent (sections 1.5.3.1, 3.2.5), has been considered as a site of ideological dilemmas. In cancer.net the mixing of the dominant metaphor themes in a single post was found to have an impact on language style, suggesting that it may be employed for specific social and psychological work, and is a particular site of change. This result warrants further investigation, for example by considering different ways of isolating and combining metaphor themes and metaphoremes, to consider more precisely the work to which the mixing of metaphor relates, both generally and within specific communities. This could be related to a more detailed investigation of metaphor as vague language that may be used to target and realise communication goals (Zhang, 2011), and, building on the insights of the current study, to a more detailed investigation of vague language used in conjunction with metaphor, with the mixing of metaphor also investigated as being potentially more vague

and more associated with use of vague language than is a single conceptual metaphor used alone.

As discussed in chapter six of this thesis, those diagnosed as autistic are around three times as likely to have suicidal ideation. Reference was made to recent deaths of autistic young adults in conjunction with suicide.net. The mother of one of those young adults stated that the warnings and support notices on suicide.net are not sufficient for vulnerable people such as her son. While the focus in the current study on more general patterns has been productive, this could be used as a platform from which to consider the experience of particular psychosocial groups in the context of the fora. For example, the issue whether particular psychosocial groups are able to manage multiple metaphorical conceptions of the same target concept simultaneously. The antithetical recontextualisation of LIFE IS A JOURNEY into the *your journey* metaphoreme and *catch the bus* metaphor was shown in this thesis to be therapeutic for participants in general in that it makes the concept of suicide accessible and supports discussion with others that have a similar stance, while also helping to characterise suicide.net as a community. However, for other participants, such as those diagnosed as autistic, the vagueness and flexibility of such recontextualisation may not be understood. The recontextualisation of LIFE IS A JOURNEY taking place at Linell's (2009) first (intra text) and second (inter text) levels may be understood at the third level of recontextualisation - the level of discourses (Foucault, 1971) - such that it becomes an overarching principle, rather than being considered contextually in relation to multiple conceptualisations. This is the case in social and cultural groups, including political and religious groups, and those designated as conspiracy theorists. For example, the QAnon conspiracy theory which led to an attack on the United States Capitol on 6th January 2021. As has been supported in the current study, people become the products of

discourse as well as the producers of it (Edley, 2001, p. 190). It seems important, then, to consider the specific experience of groups with particular psychosocial disabilities on suicide.net, and in other communities, for example by making linguistic comparisons focusing on metaphor use between communities specifically related to a particular psychosocial disability, and more general fora such as suicide.net.

Finally, although the accuracy of the supervised models in the current study was highly creditable, it would be productive to explore this methodology further. Supervised learning was used here as a method of validating the organisation of the identified metaphoric terms into themes, on the basis that if an algorithm can be trained to accurately predict the separate themes, then they are distinct, and have been correctly identified in terms of their constituent terms. Supervised learning applied in this manner was also used to identify metaphor in the cancer.net corpus, a process which supported the primary prediction of the thesis exemplified in section 1.1.2, that where metaphor is used to characterise a concept, the surrounding language would be of a style that has been associated with better mental health. The success of the supervised method in locating otherwise unidentified metaphor (section 4.3.4.2.1) may support its more general use for this purpose. However, the supervised learning in the current investigation is quite general, with a focus on two models to cover all applications, and with analysis focused primarily on the higher-level weighted average F1 score. Although a more detailed analysis is undertaken to support specific theoretical considerations, for example consideration of the impact of matched and unmatched condition sizes in section 4.3.3.2, more detailed analysis throughout could support the refinement of the models such that they may become a valuable and potentially more general method of identifying metaphor in a large text corpus, where, as discussed in chapter one, finding metaphor automatically in text is a pressing issue for many research areas. Further work could

include use of the transformer models that the current study did not have the resources to implement fully, which have proved beneficial in many applications of natural language processing.

LIST OF REFERENCES

- Althusser, L. (2001). *Lenin and Philosophy and Other Essays*. NYU Press.
<https://www.jstor.org/stable/j.ctt9qgh9v>
- AoIR. (2019). *Internet Research: Ethical Guidelines 3.0*. <https://aoir.org/reports/ethics3.pdf>
- BAAL. (2021). *Recommendations on Good Practice in Applied Linguistics. 4th Edition*.
<https://www.baal.org.uk/who-we-are/resources/BAAL-Good-Practice-Guidelines-2021.pdf>
- Bagatell, N. (2010). From Cure to Community: Transforming Notions of Autism. *Ethos*, 38(1), 33–55. <https://doi.org/10.1111/j.1548-1352.2009.01080.x>
- Baker, D., & Fortune, S. (2008). Understanding self-harm and suicide websites: A qualitative interview study of young adult website users. *Crisis : The Journal of Crisis Intervention and Suicide Prevention*, 29(3), 118–122. <https://doi.org/10.1027/0227-5910.29.3.118>
- Baldwin, M., Landau, M. J., & Swanson, T. J. (2018). Metaphors can give life meaning. *Self and Identity*, 17(2), 163–193. <https://doi.org/10.1080/15298868.2017.1368696>
- Becker, E. (1997). *The denial of death / Ernest Becker*. Free Press.
- Bell, V. (2007). Online information, extreme communities and internet therapy: Is the internet good for our mental health? *Journal of Mental Health*, 16(4), 445–457.
<https://doi.org/10.1080/09638230701482378>
- Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.
- Billig, M., Condor, S., Edwards, D., Gane, M., Middleton, D., & Radley, A. (1988). *Ideological dilemmas: A social psychology of everyday thinking*. Sage Publications, Inc.
- Birke, J., & Sarkar, A. (2006). A Clustering Approach for Nearly Unsupervised Recognition of Nonliteral Language. *11th Conference of the European Chapter of the Association for Computational Linguistics*, 329–336. <https://aclanthology.org/E06-1042>
- Black, M. (1977). More about metaphor. *Dialectica*, 31(3/4), 431–457.

- Boers, F., & Littlemore, J. (2000). Cognitive Style Variables in Participants' Explanations of Conceptual Metaphors. *Metaphor and Symbol, 15*(3), 177–187.
https://doi.org/10.1207/S15327868MS1503_4
- Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition, 75*(1), 1–28. [https://doi.org/10.1016/S0010-0277\(99\)00073-6](https://doi.org/10.1016/S0010-0277(99)00073-6)
- Bowdle, B. F., & Gentner, D. (2005). The Career of Metaphor. *Psychological Review, 112*(1), 193–216. <https://doi.org/10.1037/0033-295X.112.1.193>
- Bowker, J. (1996). Cancer, Individual Process, and Control: A Case Study in Metaphor Analysis. *Health Communication, 8*(1), 91–104. https://doi.org/10.1207/s15327027hc0801_5
- Boyd, M., Halle, M., & Sharpe, A. (2020, November 21). 'I found my 23-year-old son dead after he killed himself with £8 substance'. Mirror. <https://www.mirror.co.uk/news/uk-news/i-found-23-year-old-23046681>
- BPS. (2017). *Ethics Guidelines for Internet-mediated Research*. <https://www.bps.org.uk/news-and-policy/ethics-guidelines-internet-mediated-research>
- Brezina, V. (2018). *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge University Press.
<https://doi.org/10.1017/9781316410899>
- Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems, 30*(1–7), 107–117.
[https://doi.org/10.1016/S0169-7552\(98\)00110-X](https://doi.org/10.1016/S0169-7552(98)00110-X)
- Byrne A, Ellershaw J, Holcombe C, Salmon P, Byrne, A., Ellershaw, J., Holcombe, C., & Salmon, P. (2002). Patients' experience of cancer: Evidence of the role of 'fighting' in collusive clinical communication. In *Patient Education & Counseling* (Vol. 48, Issue 1, pp. 15–22). Elsevier BV. [https://doi.org/10.1016/s0738-3991\(02\)00094-0](https://doi.org/10.1016/s0738-3991(02)00094-0)
- Cameron, L. (2003). *Metaphor in Educational Discourse*. Bloomsbury Publishing.
- Cameron, L. (2007). Patterns of metaphor use in reconciliation talk. *Discourse & Society, 18*(2), 197–222. <https://doi.org/10.1177/0957926507073376>

- Cameron, L., & Deignan, A. (2003). Combining Large and Small Corpora to Investigate Tuning Devices Around Metaphor in Spoken Discourse. *Metaphor and Symbol*, 18(3), 149–160. https://doi.org/10.1207/S15327868MS1803_02
- Cameron, L., & Deignan, A. (2006). The Emergence of Metaphor in Discourse. *Applied Linguistics*, 27(4), 671–690. <https://doi.org/10.1093/applin/aml032>
- Cameron, L. J. (2007). Patterns of metaphor use in reconciliation talk. *Discourse & Society*, 18(2), 197–222. <https://doi.org/10.1177/0957926507073376>
- Cameron, L. J., & Stelma, J. H. (2007). Metaphor clusters in discourse. *Journal of Applied Linguistics*, 1(2), 107–136. <https://doi.org/10.1558/japl.v1i2.107>
- Cameron, L., & Maslen, R. (Eds.). (2010). *Metaphor analysis: Research practice in applied linguistics, social sciences and the humanities*. Equinox Pub.
- Cameron, L., Maslen, R., Todd, Z., Maule, J., Stratton, P., & Stanley, N. (2009). The Discourse Dynamics Approach to Metaphor and Metaphor-Led Discourse Analysis. *Metaphor and Symbol*, 24(2), 63–89. <https://doi.org/10.1080/10926480902830821>
- Cardillo, E. R., Watson, C. E., Schmidt, G. L., Kranjec, A., & Chatterjee, A. (2012). From novel to familiar: Tuning the brain for metaphors. *NeuroImage (Orlando, Fla.)*, 59(4), 3212–3221. <https://doi.org/10.1016/j.neuroimage.2011.11.079>
- Casarett, D., Pickard, A., Fishman, J. M., Alexander, S. C., Arnold, R. M., Pollak, K. I., & Tulskey, J. A. (2010). Can metaphors and analogies improve communication with seriously ill patients? *Journal of Palliative Medicine*, 13(3), 255–260. <https://doi.org/10.1089/jpm.2009.0221>
- Casasanto, D. (2013). Development of Metaphorical Thinking: The Role of Language. In *Language and the Creative Mind* (p. 16). Stanford CA: CSLI Publications.
- Casasanto, D. (2017). *The Hierarchical Structure of Mental Metaphors* (pp. 46–61). Cambridge University Press. <https://doi.org/10.1017/9781108182324.004>
- Cassidy, S., & Rodgers, J. (2017). Understanding and prevention of suicide in autism. *The Lancet Psychiatry*, 4(6), e11. [https://doi.org/10.1016/S2215-0366\(17\)30162-1](https://doi.org/10.1016/S2215-0366(17)30162-1)

- Charteris-Black, J. (2016). The 'dull roar' and the 'burning barbed wire pantyhose'. In *Mixing Metaphor*. John Benjamins Publishing. <https://benjamins.com/catalog/milcc.6.08cha>
- Chollet, F. (2018). *Deep learning with Python*. Manning Publications Co.
- Choudhury, M. D., & Kiciman, E. (2017). *The Language of Social Support in Social Media and its Effect on Suicidal Ideation Risk*. 27.
- Chung, D., Ryan, C., & Large, M. (2016). Commentary: Adverse experiences in psychiatric hospitals might be the cause of some postdischarge suicides. *Bulletin of the Menninger Clinic*, 80, 371–375. <https://doi.org/10.1521/bumc.2016.80.4.371>
- Chung, D. T., Ryan, C. J., Hadzi-Pavlovic, D., Singh, S. P., Stanton, C., & Large, M. M. (2017). Suicide Rates After Discharge From Psychiatric Facilities. *JAMA Psychiatry*, 74(7), 694–702. <https://doi.org/10.1001/jamapsychiatry.2017.1044>
- Churchland, P. (1988). *Matter and Consciousness*. The MIT Press.
<https://mitpress.mit.edu/books/matter-and-consciousness-revised-edition>
- Closure of suicide forums—Early Day Motions—UK Parliament*. (2020, November 16).
<https://edm.parliament.uk/early-day-motion/57703/closure-of-suicide-forums>
- Cooper, D. E. (1989). *Metaphor*. Blackwell.
- Coppersmith, G. A., Harman, C. T., & Dredze, M. H. (2014). *Measuring Post Traumatic Stress Disorder in Twitter*. 4.
- Crain, S. P., Zhou, K., Yang, S.-H., & Zha, H. (2012). Dimensionality Reduction and Topic Modeling: From Latent Semantic Indexing to Latent Dirichlet Allocation and Beyond. In C. C. Aggarwal & C. Zhai (Eds.), *Mining Text Data* (pp. 129–161). Springer US.
https://doi.org/10.1007/978-1-4614-3223-4_5
- Czechmeister, C. A. (1994). Metaphor in illness and nursing: A two-edged sword. A discussion of the social use of metaphor in everyday language, and implications of nursing and nursing education. *Journal of Advanced Nursing*, 19(6), 1226–1233.
<https://doi.org/10.1111/j.1365-2648.1994.tb01208.x>

- Davidson, D. (1978). What Metaphors Mean. *Critical Inquiry*, 5, 31–47.
- De Choudhury, M., Counts, S., & Horvitz, E. (2013). Predicting postpartum changes in emotion and behavior via social media. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3267–3276. <https://doi.org/10.1145/2470654.2466447>
- Deignan, A. (2005). Metaphor and Corpus Linguistics. In *Celcr.6*. John Benjamins Publishing Company. <https://benjamins.com/catalog/celcr.6>
- Deignan, A. (2011). Deliberateness is not unique to metaphor: A response to Gibbs. *Metaphor and the Social World*, 1(1), 57–60. <https://doi.org/10.1075/msw.1.1.05dei>
- Deignan, A. (2017). From linguistic to conceptual metaphors. In *The Routledge handbook of metaphor and language* (pp. 102–116). Routledge.
- Deignan, A., Semino, E., & Paul, S.-A. (2019). Metaphors of Climate Science in Three Genres: Research Articles, Educational Texts, and Secondary School Student Talk. *Applied Linguistics*, 40(2), 379–403. <https://doi.org/10.1093/applin/amx035>
- Derrida, J. (1974). White Mythology: Metaphor in the Text of Philosophy. *New Literary History*, 6(1), 5–74. <https://doi.org/10.2307/468341>
- Desmet, B., & Hoste, V. (2018). Online suicide prevention through optimised text classification. *Information Sciences*, 439–440, 61–78. <https://doi.org/10.1016/j.ins.2018.02.014>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
- Edley, N. (2001). Interpretative Repertoires, Ideological Dilemmas, and Subject Positions. In *Discourse as Data*. Sage.
- Edwards, D. (1999). Emotion Discourse. *Culture & Psychology*, 5(3), 271–291. <https://doi.org/10.1177/1354067X9953001>

- Edwards, D. (2004). Discursive Psychology. In *Handbook of Language and Social Interaction* (pp. 557–593). Taylor & Francis Group.
<http://ebookcentral.proquest.com/lib/bham/detail.action?docID=255632>
- Edwards, D. (2005). Discursive psychology. In *Handbook of language and social interaction* (pp. 257–273). Lawrence Erlbaum Associates Publishers.
- Egnoto, M. J., & Griffin, D. J. (2016). Analyzing Language in Suicide Notes and Legacy Tokens: Investigating Clues to Harm of Self and Harm to Others in Writing. *Crisis*, 37(2), 140–147.
<https://doi.org/10.1027/0227-5910/a000363>
- Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., Crutchley, P., Preotăciuc-Pietro, D., Asch, D. A., & Schwartz, H. A. (2018). Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences*, 115(44), 11203–11208.
<https://doi.org/10.1073/pnas.1802331115>
- El Refaie, E. (2015). Scoring a goal or an own-goal against disease?: A multilevel framework for describing metaphor coherence in health campaigns. *Metaphor and the Social World*, 5(1), 102–123. <https://doi.org/10.1075/msw.5.1.06ref>
- Fainsilber, L., & Ortony, A. (1987). Metaphorical Uses of Language in the Expression of Emotions. *Metaphor and Symbolic Activity*, 2(4), 239–250.
https://doi.org/10.1207/s15327868ms0204_2
- Fauconnier, G., & Turner, M. (2002). *The way we think: Conceptual blending and the mind's hidden complexities* (pp. xvii, 440). Basic Books.
- Fetterman, A. K., Bair, J. L., Werth, M., Landkammer, F., & Robinson, M. D. (2016). The Scope and Consequences of Metaphoric Thinking: Using Individual Differences in Metaphor Usage to Understand How Metaphor Functions. *Journal of Personality and Social Psychology*, 110(3), 458–476. <https://doi.org/10.1037/pspp0000067>
- Foucault, M. (1971). Orders of discourse. *Social Science Information*, 10(2), 7–30.
<https://doi.org/10.1177/053901847101000201>

- Frankl, V. E. (2004). *Man's search for meaning: The classic tribute to hope from the Holocaust / Viktor E. Frankl ; part one translated by Ilse Lasch ; preface by Gordon W. Allport*. Rider.
- Fullagar, S. (2003). Wasted lives: The social dynamics of shame and youth suicide. *Journal of Sociology*, 39(3), 291–307. <https://doi.org/10.1177/0004869003035076>
- Fuoli, M., Littlemore, J., & Turner, S. (2021). Sunken ships and screaming banshees: Metaphor and evaluation in film reviews. *English Language and Linguistics*, 1–29. <https://doi.org/10.1017/S1360674321000046>
- Gajewski, M. (2015, September 28). *May I take your metaphor? – How we talk about cancer - Cancer Research UK - Cancer news*. Cancer Research UK - Science Blog. <https://news.cancerresearchuk.org/2015/09/28/may-i-take-your-metaphor-how-we-talk-about-cancer/>
- Garrett, R. K. (2009). Echo chambers online?: Politically motivated selective exposure among Internet news users. *Journal of Computer-Mediated Communication*, 14(2), 265–285. <https://doi.org/10.1111/j.1083-6101.2009.01440.x>
- Gibbs, R. (2001). Making good psychology out of blending theory. *Cognitive Linguistics*, 11(3–4). <https://doi.org/10.1515/cogl.2001.020>
- Gibbs, R. (2017). *Metaphor Wars: Conceptual Metaphors in Human Life*. Cambridge University Press. <https://doi.org/10.1017/9781107762350>
- Gibbs, R., & Chen, E. (2017). Taking metaphor studies back to the Stone Age: A reply to Xu, Zhang, and Wu (2016). *Intercultural Pragmatics*, 14. <https://doi.org/10.1515/ip-2017-0005>
- Gibbs, R. W. (1994). *The poetics of mind: Figurative thought, language, and understanding / Raymond W. Gibbs*. University Press.
- Gibbs, R. W. (2009). Why Do Some People Dislike Conceptual Metaphor Theory? *Cognitive Semiotics*, 5(1–2), 14–36. <https://doi.org/10.1515/cogsem.2013.5.12.14>

- Gibbs, R. W. (2015). Does deliberate metaphor theory have a future? *Journal of Pragmatics*, 90, 73–76. <https://doi.org/10.1016/j.pragma.2015.03.016>
- Gibbs, R. W., & Franks, H. (2002). Embodied Metaphor In Women's Narratives About Their Experiences With Cancer. *Health Communication*, 14(2), 139–165. https://doi.org/10.1207/S15327027HC1402_1
- Gibbs, R. W., & Matlock, T. (2008). *Metaphor, imagination, and simulation* (pp. 161–176). Cambridge University Press. <https://doi.org/10.1017/CBO9780511816802.011>
- Gibs, R., & Franks, H. (2002). Embodied metaphor in women's narratives about their experiences with cancer. In *Health Communication* (Vol. 14, Issue 2, pp. 139–166). Philadelphia, Pennsylvania, Taylor & Francis Ltd. https://doi.org/10.1207/s15327027hc1402_1
- Gilat, I., & Shahar, G. (2007). Emotional First Aid for a Suicide Crisis: Comparison between Telephonic Hotline and Internet. *Psychiatry: Interpersonal and Biological Processes*, 70(1), 12–18. <https://doi.org/10.1521/psyc.2007.70.1.12>
- Goatly, A. (2011). *The language of metaphors* (2nd ed). Routledge.
- Grady, J. (1997). *Foundations of Meaning: Primary Metaphors and Primary Scenes* [University of California, Berkeley]. <https://escholarship.org/uc/item/3g9427m2>
- Grady, J. (2010). Metaphor. In *The Oxford Handbook of Cognitive Linguistics* (pp. 188–213). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199738632.001.0001>
- Grady, J., Oakley, T., & Coulson, S. (1999). Blending and metaphor. *Metaphor in Cognitive Linguistics*, 101. <https://www.jbe-platform.com/content/books/9789027284013-cilt.175.07gra>
- Gramsci, A. (2007). *Selections from the prison notebooks of Antonio Gramsci* (Reprinted). Lawrence and Wishart.
- Grohol, J. M. (1998). Future clinical directions: Professional development, pathology, and psychotherapy on-line. In *Psychology and the Internet: Intrapersonal, interpersonal, and transpersonal implications* (pp. 111–140). Academic Press.

- Gumperz, J. (1968). The Speech Community. In *Linguistic Anthropology: A reader*.
- Guntuku, S. C., Schneider, R., Pelullo, A., Young, J., Wong, V., Ungar, L., Polsky, D., Volpp, K. G., & Merchant, R. (2019). Studying expressions of loneliness in individuals using twitter: An observational study. *BMJ Open*, 9(11), e030355. <https://doi.org/10.1136/bmjopen-2019-030355>
- Harlow, L. L., Newcomb, M. D., & Bentler, P. M. (1986). Depression, self-derogation, substance use, and suicide ideation: Lack of purpose in life as a mediational factor. *Journal of Clinical Psychology*, 42(1), 5–21. [https://doi.org/10.1002/1097-4679\(198601\)42:1<5::AID-JCLP2270420102>3.0.CO;2-9](https://doi.org/10.1002/1097-4679(198601)42:1<5::AID-JCLP2270420102>3.0.CO;2-9)
- Harrow, A., Wells, M., Humphris, G., Taylor, C., & Williams, B. (2008). “Seeing is believing, and believing is seeing”: An exploration of the meaning and impact of women’s mental images of their breast cancer and their potential origins. *Patient Education and Counseling*, 73(2), 339–346. <https://doi.org/10.1016/j.pec.2008.07.014>
- Harvey, K. (2014). *Investigating adolescent health communication: A corpus linguistics approach*. Bloomsbury Academic.
- Hauser, D. J., & Schwarz, N. (2020). The War on Prevention II: Battle Metaphors Undermine Cancer Treatment and Prevention and Do Not Increase Vigilance. *Health Communication*, 35(13), 1698–1704. <https://doi.org/10.1080/10410236.2019.1663465>
- Hearst, M. A. (1992). Automatic acquisition of hyponyms from large text corpora. *Proceedings of the 14th Conference on Computational Linguistics -*, 2, 539. <https://doi.org/10.3115/992133.992154>
- Hedley, D., Uljarević, M., Wilmot, M., Richdale, A., & Dissanayake, C. (2018). Understanding depression and thoughts of self-harm in autism: A potential mechanism involving loneliness. *Research in Autism Spectrum Disorders*, 46, 1–7. <https://doi.org/10.1016/j.rasd.2017.11.003>

- Heintz, I., Gabbard, R., Srinivasan, M., Barner, D., Black, D. S., Freedman, M., & Weischedel, R. (2013). *Automatic Extraction of Linguistic Metaphor with LDA Topic Modeling*. 9.
- Hendricks, R. K., Demjén, Z., Semino, E., & Boroditsky, L. (2018). Emotional Implications of Metaphor: Consequences of Metaphor Framing for Mindset about Cancer. *Metaphor and Symbol*, 33(4), 267–279. <https://doi.org/10.1080/10926488.2018.1549835>
- Herring, S. (1999). Interactional Coherence in Cmc. *Journal of Computer-Mediated Communication*, 4(JCMC444). <https://doi.org/10.1111/j.1083-6101.1999.tb00106.x>
- Herring, S. C. (2004). Computer-Mediated Discourse Analysis: An Approach to Researching Online Behavior. In S. Barab, R. Kling, & J. H. Gray (Eds.), *Designing for Virtual Communities in the Service of Learning* (1st ed., pp. 338–376). Cambridge University Press. <https://doi.org/10.1017/CBO9780511805080.016>
- Hicks JA, Trent J, Davis WE, & King LA. (2012). Positive affect, meaning in life, and future time perspective: An application of socioemotional selectivity theory. In *Psychology & Aging* (Vol. 27, Issue 1, pp. 181–190). Washington, District of Columbia, American Psychological Association. <https://doi.org/10.1037/a0023965>
- Honnibal, M., & Montani, I. (2021a). *Linguistic Features · spaCy Usage Documentation*. Linguistic Features. <https://spacy.io/usage/linguistic-features#vectors-similarity>
- Honnibal, M., & Montani, I. (2021b). *SpaCy · Industrial-strength Natural Language Processing in Python*. SpaCy. <https://spacy.io/>
- Horne, J., & Wiggins, S. (2009). Doing being ‘on the edge’: Managing the dilemma of being authentically suicidal in an online forum. *Sociology of Health & Illness*, 31(2), 170–184. <https://doi.org/10.1111/j.1467-9566.2008.01130.x>
- Hovy, D., Shrivastava, S., Jauhar, S. K., Sachan, M., Goyal, K., Li, H., Sanders, W., & Hovy, E. (2013). *Identifying Metaphorical Word Use with Tree Kernels*. 6.
- Hunt, D., & Brookes, G. (2020). *Corpus, Discourse and Mental Health*. Bloomsbury Methuen Drama. <https://doi.org/10.5040/9781350059207>

- Iacobelli, F., Gill, A. J., Nowson, S., & Oberlander, J. (2011). Large Scale Personality Classification of Bloggers. In S. D'Mello, A. Graesser, B. Schuller, & J.-C. Martin (Eds.), *Affective Computing and Intelligent Interaction* (pp. 568–577). Springer.
https://doi.org/10.1007/978-3-642-24571-8_71
- Jaeger, J., Lindblom, K. M., Parker-Guilbert, K., & Zoellner, L. A. (2014). Trauma narratives: It's what you say, not how you say it. *Psychological Trauma: Theory, Research, Practice, and Policy*, 6(5), 473–481. <https://doi.org/10.1037/a0035239>
- Jia, L., & Smith, E. R. (2013). Distance makes the metaphor grow stronger: A psychological distance model of metaphor use. *Journal of Experimental Social Psychology*, 49(3), 492–497. <https://doi.org/10.1016/j.jesp.2013.01.009>
- Joiner, T. (2007). *Why people die by suicide* (1. Harvard University Press paperback ed). Harvard University Press.
- Jordan, J. T., & McNiel, D. E. (2020). Perceived Coercion During Admission Into Psychiatric Hospitalization Increases Risk of Suicide Attempts After Discharge. *Suicide & Life-Threatening Behavior*, 50(1), 180–188. <https://doi.org/10.1111/sltb.12560>
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, XX(1), 141–151.
- Katz, A. N. (1996). Experimental Psycholinguistics and Figurative Language: Circa 1995. *Metaphor and Symbolic Activity*, 11(1), 17–37. https://doi.org/10.1207/s15327868ms1101_2
- Kelly, B. (2019). I try to reach out beyond my own internet echo chamber. *Independent (London, England : 1986)*, 37-.
- Keras: The Python deep learning API. (2021). Keras: The Python Deep Learning API.
<https://keras.io/>
- Kessler, J. (2017). Scattertext: A Browser-Based Tool for Visualizing how Corpora Differ. *Proceedings of ACL 2017, System Demonstrations*, 85–90.
<https://www.aclweb.org/anthology/P17-4015>

- Kilgarrieff, A. (2009). Simple maths for keywords. *Proceedings of the Corpus Linguistics Conference*, 6.
- King, L. A., Hicks, J. A., Krull, J. L., & del Gaiso, A. K. (2006). Positive Affect and the Experience of Meaning in Life. *Journal of Personality & Social Psychology*, 90(1), 179–197.
<https://doi.org/10.1037/0022-3514.90.1.179>
- Kirklin D. (2007). Truth telling, autonomy and the role of metaphor. In *Journal of Medical Ethics* (Vol. 33, Issue 1, pp. 11–15). BMJ Publishing Group.
- Kirmayer, L. J., Raikhel, E., & Rahimi, S. (2013). Cultures of the Internet: Identity, community and mental health. *Transcultural Psychiatry*, 50(2), 165–191.
<https://doi.org/10.1177/1363461513490626>
- Kölves, K., Fitzgerald, C., Nordentoft, M., Wood, S. J., & Erlangsen, A. (2021). Assessment of Suicidal Behaviors Among Individuals With Autism Spectrum Disorder in Denmark. *JAMA Network Open*, 4(1), e2033565. <https://doi.org/10.1001/jamanetworkopen.2020.33565>
- Koplenig, A. (2017). Why the quantitative analysis of diachronic corpora that does not consider the temporal aspect of time-series can lead to wrong conclusions. *Digital Scholarship in the Humanities*, 32(1), 159–168. <https://doi.org/10.1093/llc/fqv030>
- Korff, J., & Biemann, T. (2020). Adjusting the lookout: Subjective health, loneliness, and life satisfaction predict future time perspective. *Psychology and Aging*, 35(8), 1170–1183.
<https://doi.org/10.1037/pag0000513>
- Kovecses, Z. (2000). *Metaphor and Emotion*. Cambridge University Press.
- Kövecses, Z. (2008). Metaphor and Emotion. In *The Cambridge Handbook of Metaphor and Thought* (pp. 380–396).
- Kövecses, Z. (2015). *Where Metaphors Come From: Reconsidering Context in Metaphor*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780190224868.001.0001>
- Kövecses, Z. (2016). A view of “mixed metaphor” within a conceptual metaphor theory framework. In R. W. Gibbs (Ed.), *Metaphor in Language, Cognition, and Communication*

- (Vol. 6, pp. 3–16). John Benjamins Publishing Company.
- <https://doi.org/10.1075/milcc.6.01kov>
- Krishnakumaran, S., & Zhu, X. (2007). Hunting elusive metaphors using lexical resources. *Proceedings of the Workshop on Computational Approaches to Figurative Language - FigLanguages '07*, 13–20. <https://doi.org/10.3115/1611528.1611531>
- Lakoff, G., Espenson, J., & Schwartz, A. (1991). *Master Metaphor List*.
<http://araw.mede.uic.edu/~alansz/metaphor/METAPHORLIST.pdf>
- Lakoff, G., & Johnson, M. (2003). *Metaphors we live by*. University of Chicago Press.
- Lanceley, A., & Clark, J. M. (2013). Cancer in Other Words? The Role of Metaphor in Emotion Disclosure in Cancer Patients: Cancer in Other Words? *British Journal of Psychotherapy*, 29(2), 182–201. <https://doi.org/10.1111/bjp.12023>
- Landau, M. J. (2017). *Conceptual metaphor in social psychology: The poetics of everyday life* (1st Edition). Routledge, Taylor & Francis Group.
- Lavis, A., & Winter, R. (2020). #Online harms or benefits? An ethnographic analysis of the positives and negatives of peer-support around self-harm on social media.
[https://research.birmingham.ac.uk/portal/en/publications/online-harms-or-benefits-an-ethnographic-analysis-of-the-positives-and-negatives-of-peersupport-around-selfharm-on-social-media\(297e63cf-e357-4a09-b4b3-ac51091db945\).html](https://research.birmingham.ac.uk/portal/en/publications/online-harms-or-benefits-an-ethnographic-analysis-of-the-positives-and-negatives-of-peersupport-around-selfharm-on-social-media(297e63cf-e357-4a09-b4b3-ac51091db945).html)
- Leezenberg, M. (2009). From Cognitive Linguistics to Social Science: Thirty Years after Metaphors We Live By. *Cognitive Semiotics*, 5(1–2), 140–152.
<https://doi.org/10.1515/cogsem.2013.5.12.140>
- Leong, C. W. (Ben), Beigman Klebanov, B., Hamill, C., Stemle, E., Ubale, R., & Chen, X. (2020). A Report on the 2020 VUA and TOEFL Metaphor Detection Shared Task. *Proceedings of the Second Workshop on Figurative Language Processing*, 18–29.
<https://doi.org/10.18653/v1/2020.figlang-1.3>

- Li, H., Zhu, K. Q., & Wang, H. (2013). Data-Driven Metaphor Recognition and Explanation. *Transactions of the Association for Computational Linguistics*, 1, 379–390.
https://doi.org/10.1162/tacl_a_00235
- Li, L., & Sporleder, C. (2009). *Classifier combination for contextual idiom detection without labelled data*. 9.
- Li, Q. (2020). Text-based Computer-mediated Discourse Analysis: What Causes an Online Group to Become a Virtual Community? *International Journal of Social Science Studies*, 8(4), 27.
<https://doi.org/10.11114/ijsss.v8i4.4867>
- Linell, P. (2009). *Rethinking language, mind, and world dialogically: Interactional and contextual theories of human sense-making*. Information Age Pub.
- Lingis, A. (1994). *The community of those who have nothing in common / Alphonso Lingis*. Indiana University Press.
- Littlemore, J. (2017). *Metonymy*.
[https://research.birmingham.ac.uk/portal/en/publications/metonymy\(94155423-ecc9-419e-8c95-648c9202c987\).html](https://research.birmingham.ac.uk/portal/en/publications/metonymy(94155423-ecc9-419e-8c95-648c9202c987).html)
- Littlemore, J. (2019). *Metaphors in the mind: Sources of variation in embodied metaphor / Jeannette Littlemore*. Cambridge University Press.
- Low, G., Todd, Z., Deignan, A., & Cameron, L. (Eds.). (2010). *Researching and Applying Metaphor in the Real World* (Vol. 26). John Benjamins Publishing Company.
<https://doi.org/10.1075/hcp.26>
- Lyons, M., Aksayli, N. D., & Brewer, G. (2018). Mental distress and language use: Linguistic analysis of discussion forum posts. *Computers in Human Behavior*, 87, 207–211.
<https://doi.org/10.1016/j.chb.2018.05.035>
- Marchant, A., Hawton, K., Stewart, A., Montgomery, P., Singaravelu, V., Lloyd, K., Purdy, N., Daine, K., & John, A. (2017). *A systematic review of the relationship between internet use, self-harm and suicidal behaviour in young people::The good, the bad and the*

- unknown. [https://research.birmingham.ac.uk/portal/en/publications/a-systematic-review-of-the-relationship-between-internet-use-selfharm-and-suicidal-behaviour-in-young-people\(dc610242-b267-400d-920e-1bd57332ecb0\).html](https://research.birmingham.ac.uk/portal/en/publications/a-systematic-review-of-the-relationship-between-internet-use-selfharm-and-suicidal-behaviour-in-young-people(dc610242-b267-400d-920e-1bd57332ecb0).html)
- Marková, I. (Ed.). (2007). *Dialogue in focus groups: Exploring socially shared knowledge*. Equinox.
- Messner, B. A., & Buckrop, J. J. (2000). Restoring order: Interpreting suicide through a Burkean lens. *Communication Quarterly*, 48(1), 1–18.
<https://doi.org/10.1080/01463370009385575>
- Mihalcea, R., & Tarau, P. (2004). *TextRank: Bringing Order into Texts*. 8.
- Mohammad, S. (2021). *NRC Emotion Lexicon*. <https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>
- Mohammad, S., Shutova, E., & Turney, P. (2016). Metaphor as a Medium for Emotion: An Empirical Study. *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*, 23–33. <https://doi.org/10.18653/v1/S16-2003>
- Mohler, M., Bracewell, D., Tomlinson, M., & Hinote, D. (2013). *Semantic Signatures for Example-Based Linguistic Metaphor Detection*. 9.
- Morgan, M. H. (2014). *Speech communities*. Cambridge University Press.
- Moskaluk, K. (2020). *Dizziness of Freedom: The influence of maladaptive anxiety on metaphorical meaning-making and the Motivation & Sedimentation Model (MSM)*.
<http://lup.lub.lu.se/student-papers/record/9033994>
- Mother speaks out against suicide forum after son's death. (2021, May 8). *BBC News*.
<https://www.bbc.com/news/uk-northern-ireland-56845565>
- Müller, A. C., & Guido, S. (2016). *Introduction to machine learning with Python: A guide for data scientists* (First edition). O'Reilly Media, Inc.
- Müller, C. (2008). *Metaphors Dead and Alive, Sleeping and Waking: A Dynamic View*.
<https://press.uchicago.edu/ucp/books/book/chicago/M/bo5877524.html>

- Müller, C. (2010). Metaphors Dead and Alive, Sleeping and Waking: A Dynamic View. *Metaphor and Symbol*, 25(2), 114–121. <https://doi.org/10.1080/10926481003716026>
- Müller, C. (2011). Are ‘deliberate’ metaphors really special?: Deliberateness in the light of metaphor activation. *Metaphor and the Social World*, 1(1), 61–66. <https://doi.org/10.1075/msw.1.1.06mul>
- Müller, C. (2016). Why mixed metaphors make sense. In *Mixing Metaphor* (Vol. 6, pp. 31-56 <https://doi.org/10.1075/milcc.6.03mul>). John Benjamins Publishing Company.
- Murakami, A., Thompson, P., Hunston, S., & Vajn, D. (2017). ‘What is this corpus about?’: Using topic modelling to explore a specialised corpus. *Corpora*, 12(2), 243–277. <https://doi.org/10.3366/cor.2017.0118>
- Nathan, P., Ankush Chander, Guitton, L., Michał Nowotka, Gupta, H., Carter, J., Chen, J., Thejeshwar, S., Anna-Droid-Beep, Athanasakis, D., Katharine, & æriess. (2021). *DerwenAI/pytextrank: V3.1.0 release on PyPi (v3.1.0)* [Computer software]. Zenodo. <https://doi.org/10.5281/ZENODO.4602393>
- Neuman, Y., Assaf, D., Cohen, Y., Last, M., Argamon, S., Howard, N., & Frieder, O. (2013). Metaphor Identification in Large Texts Corpora. *PLoS ONE*, 8(4), e62343. <https://doi.org/10.1371/journal.pone.0062343>
- Neuman, Y., Cohen, Y., Assaf, D., & Kedma, G. (2012). Proactive screening for depression through metaphorical and automatic text analysis. *Artificial Intelligence in Medicine*, 56(1), 19–25. <https://doi.org/10.1016/j.artmed.2012.06.001>
- Nitschke, P., & Stewart, F. (2021). *The Peaceful Pill Handbook*. <https://www.peacefulpillhandbook.com/>
- O’Connor, R. (2021). *When it is Darkest: Why people die by suicide and what we can do to help*. VERMILION.

- O'Dea, B., Larsen, M. E., Batterham, P. J., Caelear, A. L., & Christensen, H. (2017). A Linguistic Analysis of Suicide-Related Twitter Posts. *Crisis*, 38(5), 319–329.
<https://doi.org/10.1027/0227-5910/a000443>
- O'Hara, K., & Stevens, D. (2015). Echo Chambers and Online Radicalism: Assessing the Internet's Complicity in Violent Extremism. *Policy and Internet*, 7(4), 401–422.
<https://doi.org/10.1002/poi3.88>
- ONS. (2019). *Suicides in England and Wales—Office for National Statistics*.
<https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/suicidesintheunitedkingdom/2019registrations>
- Pająk, K., & Trzebiński, J. (2014). Escaping the World: Linguistic Indicators of Suicide Attempts in Poets. *Journal of Loss and Trauma*, 19(5), 389–402.
<https://doi.org/10.1080/15325024.2013.794663>
- Park, J., Ayduk, Ö., & Kross, E. (2016). Stepping back to move forward: Expressive writing promotes self-distancing. *Emotion (Washington, D.C.)*, 16(3), 349–364.
<https://doi.org/10.1037/emo0000121>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2018). Scikit-learn: Machine Learning in Python. *ArXiv:1201.0490 [Cs]*. <http://arxiv.org/abs/1201.0490>
- Pennebaker, J. (2011). *The Secret Life of Pronouns*. Bloomsbury Press.
- Pennebaker, J., Boyd, R., Jordan, K., & Blackburn, K. (2015). *The Development and Psychometric Properties of LIWC2015*. <https://doi.org/10.15781/T29G6Z>
- Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When Small Words Foretell Academic Success: The Case of College Admissions Essays. *PLOS ONE*, 9(12), e115844. <https://doi.org/10.1371/journal.pone.0115844>

- Philip, G. (2010). Metaphorical keyness in specialised corpora. In M. Bondi & M. Scott (Eds.), *Studies in Corpus Linguistics* (Vol. 41, pp. 185–204). John Benjamins Publishing Company.
<https://doi.org/10.1075/scl.41.13phi>
- Pierre, J. (2015). Culturally sanctioned suicide: Euthanasia, seppuku, and terrorist martyrdom. *World Journal of Psychiatry*, 5, 4–14. <https://doi.org/10.5498/wjp.v5.i1.4>
- Plath, S. (2021, May 13). *Edge* (<https://www.poetryfoundation.org/>) [Text/html]. Poetry Foundation; Poetry Foundation.
<https://www.poetryfoundation.org/poems/49009/edge-56d22ab50bbc1>
- Potter, J., & Wetherell, W. (1987). *Discourse and Social Psychology*. SAGE Publications Inc.
<https://us.sagepub.com/en-us/nam/discourse-and-social-psychology/book202303>
- Potts, A., & Semino, E. (2019). Cancer as a Metaphor. *Metaphor and Symbol*, 34(2), 81–95.
<https://doi.org/10.1080/10926488.2019.1611723>
- Pragglejaz. (2007). MIP: A Method for Identifying Metaphorically Used Words in Discourse. *Metaphor and Symbol*, 22(1), 1–39. <https://doi.org/10.1080/10926480709336752>
- Prandi, M. (2012). A Plea for Living Metaphors: Conflictual Metaphors and Metaphorical Swarms. *Metaphor and Symbol*, 27(2), 148–170.
<https://doi.org/10.1080/10926488.2012.667690>
- Rayson, P., Archer, D., Piao, S., & McEnery, A. M. (2004). *The UCREL semantic analysis system*. (pp. 7–12). <https://eprints.lancs.ac.uk/id/eprint/1783/>
- Reddy, M. (1979). The Conduit Metaphor A Case of Frame Conflict.pdf. *Metaphor and Thought*, 2, 284–324.
- Reijnierse, W. G., Burgers, C., Krennmayr, T., & Steen, G. J. (2018). DMIP: A Method for Identifying Potentially Deliberate Metaphor in Language Use. *Corpus Pragmatics*, 2(2), 129–147. <https://doi.org/10.1007/s41701-017-0026-7>
- Reisfield, G. M., & Wilson, G. R. (2004). Use of Metaphor in the Discourse on Cancer. *Journal of Clinical Oncology*, 22(19), 4024–4027. <https://doi.org/10.1200/JCO.2004.03.136>

- Resnik, P., Garron, A., & Resnik, R. (2013). *Using Topic Modeling to Improve Prediction of Neuroticism and Depression in College Students*. 6.
- Richa, S., Fahed, M., Khoury, E., & Mishara, B. (2014). Suicide in Autism Spectrum Disorders. *Archives of Suicide Research*, 18(4), 327–339.
<https://doi.org/10.1080/13811118.2013.824834>
- Rorty, R. (1989). *Contingency, irony and solidarity / Richard Rorty*. University Press.
- Routledge, C., Ostafin, B., Juhl, J., Sedikides, C., Cathey, C., & Liao, J. (2010). Adjusting to Death: The Effects of Mortality Salience and Self-Esteem on Psychological Well-Being, Growth Motivation, and Maladaptive Behavior. *Journal of Personality and Social Psychology*, 99(6), 897–916. <https://doi.org/10.1037/a0021431>
- r/PsychiatricFreedom—Sanctioned suicide banned. (2018). Reddit.
https://www.reddit.com/r/PsychiatricFreedom/comments/84fd21/sanctioned_suicide_banned/
- Rude, S., Gortner, E.-M., & Pennebaker, J. (2004). Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, 18(8), 1121–1133.
<https://doi.org/10.1080/02699930441000030>
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, 50(4), 696–735.
<https://doi.org/10.2307/412243>
- Salem, D. A., Bogat, G. A., & Reid, C. (1997). Mutual help goes on-line. *Journal of Community Psychology*, 25(2), 189–207. [https://doi.org/10.1002/\(SICI\)1520-6629\(199703\)25:2<189::AID-JCOP7>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1520-6629(199703)25:2<189::AID-JCOP7>3.0.CO;2-T)
- Samaritans. (2020). *Managing self-harm and suicide content online*.
- Samaritans. (2021). *The internet and suicide*. Samaritans. <https://www.samaritans.org/about-samaritans/research-policy/internet-suicide/>

- Schaub, M., & Morisi, D. (2020). Voter mobilisation in the echo chamber: Broadband internet and the rise of populism in Europe. *European Journal of Political Research*, 59(4), 752–773. <https://doi.org/10.1111/1475-6765.12373>
- Schoene, A. M., Turner, A., De Mel, G. R., & Dethlefs, N. (2021). Hierarchical Multiscale Recurrent Neural Networks for Detecting Suicide Notes. *IEEE Transactions on Affective Computing*, 1–1. <https://doi.org/10.1109/TAFFC.2021.3057105>
- Scikit-learn.org. (2021). <https://scikit-learn.org/stable/modules/decomposition.html#decompositions>
- Scrapy. (2021). *Scrapy | A Fast and Powerful Scraping and Web Crawling Framework*. Scrapy. <https://scrapy.org/>
- Semino, E. (2011). Metaphor, creativity, and the experience of pain across genres. In *Creativity, Language, Literature: The State of the Art* (pp. 83–102). Palgrave Macmillan.
- Semino, E. (2019). *A Metaphor Menu for people living with Cancer*. <http://wp.lancs.ac.uk/melc/the-metaphor-menu/>
- Semino, E., Deignan, A., & Littlemore, J. (2013). Metaphor, Genre, and Recontextualization. *Metaphor and Symbol*, 28(1), 41–59. <https://doi.org/10.1080/10926488.2013.742842>
- Semino, E., & Demjén, Z. (2017). The Cancer Card: Metaphor, Intimacy, and Humor in Online Interactions about the Experience of Cancer. In B. Hampe (Ed.), *Metaphor* (pp. 181–199). Cambridge University Press. <https://doi.org/10.1017/9781108182324.011>
- Semino, E., Demjén, Z., Hardie, A., Payne, S., & Rayson, P. (2017). *Metaphor, Cancer and the End of Life: A Corpus-Based Study* (1st ed.). Routledge. <https://doi.org/10.4324/9781315629834>
- Seraj, S., Blackburn, K. G., & Pennebaker, J. W. (2021). Language left behind on social media exposes the emotional and cognitive costs of a romantic breakup. *Proceedings of the National Academy of Sciences*, 118(7), e2017154118. <https://doi.org/10.1073/pnas.2017154118>

- Sexton, J., & Pennebaker, J. W. (2009). The Healing Powers of Expressive Writing. In *The Psychology of Creative Writing*. <https://www-cambridge-org.ezproxyd.bham.ac.uk/core/books/psychology-of-creative-writing/healing-powers-of-expressive-writing/B1BF63122E931FA980C7ED20F0E56F77>
- Shutova, E. (2015). Design and Evaluation of Metaphor Processing Systems. *Computational Linguistics*, 41(4), 579–623. https://doi.org/10.1162/COLI_a_00233
- Shutova, E., & Sun, L. (2013). Unsupervised Metaphor Identification Using Hierarchical Graph Factorization Clustering. *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 978–988. <https://www.aclweb.org/anthology/N13-1118>
- Sievert, C., & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, 63–70. <https://doi.org/10.3115/v1/W14-3110>
- sklearn.metrics.f1_score—Scikit-learn 0.24.2 documentation*. (2021). https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html
- Skorczynska, H., & Ahrens, K. (2015). A corpus-based study of metaphor signaling variations in three genres. *Text & Talk*, 35(3). <https://doi.org/10.1515/text-2015-0007>
- Skott C. (2002). Expressive metaphors in cancer narratives. In *Cancer Nursing* (Vol. 25, Issue 3, pp. 230–236). Baltimore, Maryland, Lippincott Williams & Wilkins.
- Sontag, S. (1978). *Illness as metaphor*. Farrar, Straus and Giroux.
- Steen, G. (1999). From linguistic to conceptual metaphor in five steps. In *Metaphor in Cognitive Linguistics*. John Benjamins Publishing Company.
- Steen, G. (2008). The Paradox of Metaphor: Why We Need a Three-Dimensional Model of Metaphor. *Metaphor and Symbol*, 23(4), 213–241. <https://doi.org/10.1080/10926480802426753>

- Steen, G. (2016). *Mixed metaphor is a question of deliberateness* (pp. 113–132).
<https://doi.org/10.1075/milcc.6.06ste>
- Steen, G. (2017). Deliberate Metaphor Theory: Basic assumptions, main tenets, urgent issues.
Intercultural Pragmatics, 14(1), 1–24. <https://doi.org/10.1515/ip-2017-0001>
- Steen, G. J. (2011). What does ‘really deliberate’ really mean?: More thoughts on metaphor and consciousness. *Metaphor and the Social World*, 1(1), 53–56.
<https://doi.org/10.1075/msw.1.1.04ste>
- Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A., Krennmayr, T., & Pasma, T. (2010). *A Method for Linguistic Metaphor Identification: From MIP to MIPVU* (Vol. 14). John Benjamins Publishing Company. <https://doi.org/10.1075/celcr.14>
- Stich, S. (1983). *From Folk Psychology to Cognitive Science* | The MIT Press. The MIT Press.
<https://mitpress.mit.edu/books/folk-psychology-cognitive-science>
- Stommel, W. (2009). *Entering an online support group on eating disorders: A discourse analysis* / Wyke Stommel. Rodopi.
<https://ebookcentral.proquest.com/lib/bham/detail.action?docID=556652>
- Suler, J. (2005). The online disinhibition effect. *International Journal of Applied Psychoanalytic Studies*, 2(2), 184–188. <https://doi.org/10.1002/aps.42>
- Takami, K. (2009). *Battle royale* (2nd ed). Haikasoru.
- Talmy, L. (1988). Force Dynamics in Language and Cognition. *Cognitive Science*, 12(1), 49–100.
https://doi.org/10.1207/s15516709cog1201_2
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, 29(1), 24–54. <https://doi.org/10.1177/0261927X09351676>
- Tay, D. (2011a). THERAPY IS A JOURNEY as a discourse metaphor. *Discourse Studies*, 13(1), 47–68. <https://doi.org/10.1177/1461445610387736>

- Tay, D. (2011b). Discourse markers as metaphor signalling devices in psychotherapeutic talk. *Language & Communication*, 31(4), 310–317.
<https://doi.org/10.1016/j.langcom.2011.02.001>
- Tay, D. (2019). *Time Series Analysis of Discourse: Method and Case Studies*. Routledge.
- Tay, D. (2020). Surveying views of metaphor vs. literal language in psychotherapy: A factor analysis. *Metaphor and the Social World*, 10(2), 273–291.
<https://doi.org/10.1075/msw.00007.tay>
- Thibodeau, P., & Durgin, F. H. (2008). Productive figurative communication: Conventional metaphors facilitate the comprehension of related novel metaphors. *Journal of Memory and Language*, 58(2), 521–540. <https://doi.org/10.1016/j.jml.2007.05.001>
- Trope, Y., & Liberman, N. (2003). Temporal Construal. *Psychological Review*, 110(3), 403–421.
<https://doi.org/10.1037/0033-295X.110.3.403>
- Tsvetkov, Y., Mukomel, E., & Gershman, A. (2013). *Cross-Lingual Metaphor Detection Using Common Semantic Features*. 7.
- Turney, P., Neuman, Y., Assaf, D., & Cohen, Y. (2011). Literal and Metaphorical Sense Identification through Concrete and Abstract Context. *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, 680–690.
<https://www.aclweb.org/anthology/D11-1063>
- United Nations High Commissioner for Human Rights. (2017). *Mental health and human rights*.
- Vaccari, C. (2013). From echo chamber to persuasive device? Rethinking the role of the Internet in campaigns. *New Media & Society*, 15(1), 109–127.
<https://doi.org/10.1177/1461444812457336>
- Van den Nest, M., Till, B., & Niederkrotenthaler, T. (2019). Comparing Indicators of Suicidality Among Users in Different Types of Nonprofessional Suicide Message Boards: A Linguistic Analysis. *Crisis*, 40(2), 125–133. <https://doi.org/10.1027/0227-5910/a000540>

- Veale, T., Shutova, E., & Klebanov, B. B. (2016). Metaphor: A Computational Perspective. *Synthesis Lectures on Human Language Technologies*, 9(1), 1–160.
<https://doi.org/10.2200/S00694ED1V01Y201601HLT031>
- Wallace, D. F. (2014). *Infinite jest* (Reprint). Abacus.
- Weintraub, W. (1981). *Verbal Behavior: Adaptation and Psychopathology*. Springer.
- White, M., & Dorman, S. M. (2001). Receiving social support online: Implications for health education. *Health Education Research*, 16(6), 693–707.
<https://doi.org/10.1093/her/16.6.693>
- Wiggins, S., & Potter, J. (2017). Discursive Psychology. In C. Willig & W. S. Rogers, *The SAGE Handbook of Qualitative Research in Psychology* (pp. 93–109). SAGE Publications Ltd.
<https://doi.org/10.4135/9781526405555.n6>
- Wilks, Y., Dalton, A., Allen, J., & Galescu, L. (2013). Automatic Metaphor Detection using Large-Scale Lexical Resources and Conventional Metaphor Extraction. *Proceedings of the First Workshop on Metaphor in NLP*, 36–44. <https://www.aclweb.org/anthology/W13-0905>
- Williams Camus, J. T. (2009). Metaphors of cancer in scientific popularization articles in the British press. *Discourse Studies*, 11(4), 465–495.
<https://doi.org/10.1177/1461445609105220>
- Wiltsey Stirman, S., & Pennebaker, J. W. (2001). Word Use in the Poetry of Suicidal and Nonsuicidal Poets: *Psychosomatic Medicine*, 63(4), 517–522.
<https://doi.org/10.1097/00006842-200107000-00001>
- ‘Without suicide forums, Callie might still be alive’. (2020, February 10). *BBC News*.
<https://www.bbc.com/news/newsbeat-51262240>
- Yalom, I. D. (1980). *Existential psychotherapy / Irvin D. Yalom*. Basic Books.
- Yates, A., Cohan, A., & Goharian, N. (2017). Depression and Self-Harm Risk Assessment in Online Forums. *ArXiv:1709.01848 [Cs]*. <http://arxiv.org/abs/1709.01848>
- Yates, S., Taylor, S., & Wetherell, M. (Eds.). (2001). *Discourse as data: A guide for analysis*. SAGE.

Zhang, G. (2011). *Elasticity of vague language*. 8(4), 571–599.

<https://doi.org/10.1515/iprg.2011.026>

Zhang, G. (2013). The impact of touchy topics on vague language use. *Journal of Asian Pacific*

Communication, 23(1), 87–118. <https://doi.org/10.1075/japc.23.1.06zha>

Zhong, C.-B., & Liljenquist, K. (2006). Washing Away Your Sins: Threatened Morality and Physical Cleansing. *Science (New York, N.Y.)*, 313, 1451–1452.

<https://doi.org/10.1126/science.1130726>

Zlatev, J., Jacobsson, G., & Paju, L. (2021). Desiderata for metaphor theory, the Motivation & Sedimentation Model and motion-emotion metaphoremes. In A. Soares da Silva (Ed.),

Figurative Thought and Language (Vol. 11, pp. 41–74). John Benjamins Publishing

Company. <https://doi.org/10.1075/ftl.11.02zla>