University of Huddersfield

University of Huddersfield Repository

Price, Hazel

The Discursive Construction of Mental Illness in UK Newspapers (1984-2014): A Critical Corpus Linguistic Analysis

Original Citation

Price, Hazel (2019) The Discursive Construction of Mental Illness in UK Newspapers (1984-2014): A Critical Corpus Linguistic Analysis. Doctoral thesis, University of Huddersfield.

This version is available at http://eprints.hud.ac.uk/id/eprint/35163/

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: E.mailbox@hud.ac.uk.

http://eprints.hud.ac.uk/

University of HUDDERSFIELD
Inspiring tomorrow's professionals

The Discursive Construction of Mental Illness in UK Newspapers (1984-2014): A Critical Corpus Linguistic Analysis

Hazel Price
BA (York St John) MA (Huddersfield)

A thesis submitted to the Department of Linguistics and Modern Languages in fulfilment of the requirements of the degree of Doctor of Philosophy

September 2019

To my mum

ABSTRACT

This thesis explores the representation of mental illness in the UK press. Specifically, it addresses the following central research questions: 'What do the terms 'mental health' and 'mental illness' refer to, 'How are people with mental illness named and referred to in reports on mental illness', 'What are the salient transitivity processes in news reports on mental illness' and 'Do press reports on mental illness accurately portray the symptoms of specific mental illnesses?' In order to investigate these questions, I designed and constructed the MI 1984-2014 corpus, which comprises 50,972,932 words of UK local and national news articles from 1984-2014. The stretch of time covered by the corpus is an important period for legislation related to mental health, including the 1983 Mental Health Act and the amendments to this act in 2007. I use frameworks drawn from corpus linguistics (e.g. keyness and collocation analysis) and Critical Discourse Analysis (e.g. naming and transitivity analysis) to analyse the MI 1984-2014 corpus.

The main findings from my study are as follows: (i) lexical items in the semantic domain of mental health and illness are undergoing semantic change (e.g. the term 'mental health' is being used more frequently to refer to states of mental illness via a process of socially motivated and euphemistic language change); (ii) with regard to naming practices and, in particular, naming practices that anti-stigma initiatives have identified as problematic and stigmatising, the press use identity-first forms (identified as stigmatising by mental health advocates) to refer to people with mental illness (e.g. 'a schizophrenic') more often than person-first forms (such as 'a person with schizophrenia'); despite this, early evidence suggests the press are increasingly adopting person-first language, which is the linguistic structure promoted by mental health advocates; (iii) with reference to transitivity, whilst the press overall represent the process of having mental illness as 'suffering', first-person accounts from people with mental illness are proportionally 4-times more likely to refer to having mental illness as 'experiencing' it (e.g. "I was experiencing psychosis"). I found that, overall, reports that include symptoms of mental illness are inaccurate, or are reported in contexts that are too specific to serve the purpose of properly informing the public about mental illness.

On the basis of these findings, I argue that it would be beneficial for journalists and mental health charities to make a number of changes to the way they write about mental health. One basic but important change for mental health charities would be to take account of linguistic evidence prior to creating guidelines stipulating

prescribed linguistic forms for discussing mental illness in the press. A further important change for journalists would be to more accurately depict the symptoms of mental illness in news articles and ensure that symptoms are contextualised appropriately (e.g. not used in reference to violent attacks).

This thesis is offered as a contribution to the developing field of medical humanities. It provides findings and methods for examining further the issue of the press representation of mental illness and the related impact on society (and on individuals in society) that this can have.

ACKNOWLEDGMENTS

There are very many people I need to thank for their support and encouragement during my time studying for this PhD.

Firstly, I'd like to thank everyone in linguistics at the University of Huddersfield, who have made my time writing this thesis so enjoyable, rewarding and fun. I have been lucky enough to work in a department of people during my time as a PhD student who have supported and inspired me and who have made me laugh every day. I know I will reflect on my time writing this thesis with fond memories because of the supportive environment they have created. I'd also like to give special thanks to my Huddersfield colleague, Dr Hugo Sanjurjo-González, for all his help in the process of designing and writing the script that I used in this thesis to create the subcorpora for the MI 1984-2014 corpus. I feel incredibly lucky to have been part of the linguistics team at Huddersfield.

Secondly, I'd like to thank my friends and family for their encouragement during my time writing this PhD. I'd like to thank Jack especially for his love, support and patience (and most importantly for always making the brews). I'm also incredibly grateful to my brilliant mum who will probably never read this thesis, but to whom it is dedicated. I'd like to thank her for always being there and for always teaching me the value of education.

Finally, I am hugely indebted to my supervisor, Professor Dan McIntyre, for his unfailing encouragement, support and kindness. I owe Dan thanks for so many things that whatever I write here will not accurately convey my gratitude to him. As a result of Dan's supervision and mentorship, I feel that I am not only a better linguist but that I am prepared for academia. Through his supervision, Dan has shown me what it means to be a good mentor. I am truly lucky to have Dan as a supervisor, mentor and friend.

i. The author of this thesis (including any appendices and/ or schedules to this thesis) owns any copyright in it (the "Copyright") and s/he has given The University of Huddersfield the right to use such Copyright for any administrative, promotional, educational and/or teaching purposes.

ii. Copies of this thesis, either in full or in extracts, may be made only in accordance with the regulations of the University Library. Details of these regulations may be obtained from the Librarian. Details of these regulations may be obtained from the Librarian. This page must form part of any such copies made.

iii. The ownership of any patents, designs, trademarks and any and all other intellectual property rights except for the Copyright (the "Intellectual Property Rights") and any reproductions of copyright works, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property Rights and Reproductions cannot and must not be made available for use without permission of the owner(s) of the relevant Intellectual Property Rights and/or Reproductions.

TABLE OF CONTENTS

LIST OF FIGURES **10**

LIST OF TABLES **11**

1. INTRODUCTION **14**

1.1 THE LANGUAGE OF MENTAL ILLNESS **19**
1.2 RESEARCH QUESTIONS **21**
1.3 THE MENTAL ILLNESS CORPUS 1984-2014 ..... **22**
1.4 STRUCTURE OF THIS THESIS **24**
1.5 CONVENTIONS **28**
1.6 CONCLUSION **28**

2. THE LANGUAGE OF MENTAL ILLNESS **30**

2.1 INTRODUCTION **30**
2.2 SOCIAL CONSTRUCTIONISM **31**
2.3 SOCIAL CONSTRUCTIONISM AND CDA **34**
2.4 THEMES IN RESEARCH ON MENTAL HEALTH AND LANGUAGE **39**
2.4.1 Stigma and self stigma *39*
2.4.2 Stigma reduction efforts *43*
2.5 PRESS DATA **48**
2.5.1 Mental illness in the news *49*
2.5.2 Depictions of criminality and violence *55*
2.5.3 Gender and mental illness *58*
2.5.4 Depictions of schizophrenia *60*
2.6 EXISTING RESEARCH: AN OVERVIEW **70**
2.7 THE PLACE OF THE CURRENT STUDY **71**
2.8 CONCLUSION **72**

3. ANALYTICAL METHODS 1: CORPUS LINGUISTICS **74**

3.1. INTRODUCTION **74**
3.2. CORPUS LINGUISTICS **74**
3.3. CORPUS METHODS **79**
3.3.1. Frequency analysis *79*
3.3.2. Keyness analysis *84*
3.3.3. Keyword analysis *85*
3.3.4. Key semantic domain analysis *88*
3.3.5. Collocation analysis *90*
3.3.6. N-grams *97*
3.3.7. Concordance analysis *98*
3.3.8. Semantic preference and semantic prosody *102*
3.4. CONCLUSION **105**

4. ANALYTICAL METHODS 2: CRITICAL DISCOURSE ANALYSIS **106**

4.1. CRITICAL DISCOURSE ANALYSIS **106**
4.2. TRANSITIVITY ANALYSIS **118**
4.3. NAMING ANALYSIS **128**
4.4. CONCLUSION **134**

5. CORPUS CONSTRUCTION 135

5.1. SEARCH TERMS: INTERPRETATIVELY NEUTRAL? 135
5.2. SEARCH TERMS AND SAMPLING FRAME FOR THE MI 1984-2014 CORPUS 139
5.2.1. Generating search terms 139
5.2.2. Mind 140
5.2.3. Using the Mind 'A-Z of Mental Health' to build search terms 141
5.2.4. Final search terms 145
5.3. RATIONALE FOR DATES COVERED 147
5.3.1. LexisNexis 151
5.3.2. Sampling frame for MI 1984-2014 154
5.4. DATA CLEANING 157
5.5. CONSTRUCTION OF SUBCORPORA 159
5.6. PRACTICAL ISSUES: LINGUISTICS, PROGRAMMING AND THE NATURE OF LANGUAGE 168
5.7. THE MI 1984-2014 CORPUS: ASSESSING RELEVANCE AND DISTRIBUTION OF TERMS . 170
5.8. CAVEATS OF THE CORPUS CONSTRUCTION PROCEDURE 174
5.9. CONCLUSION 175

6. THE SHIFTING MEANING OF MENTAL HEALTH AND MENTAL ILLNESS 177

6.1. EXPLORING *MENTAL ILLNESS* AND *MENTAL HEALTH* IN THE MI 1984-2014 CORPUS 181
6.2. MENTAL HEALTH 'PROBLEMS', 'CONDITIONS' AND 'ISSUES' 184
6.3. THE RISE OF 'WELLBEING' 192
6.4. ASSESSING SIMILARITY AND DIFFERENCE THROUGH COLLOCATION 199
6.5. THE COLLOCATIONAL CONTEXT OF 'MENTAL HEALTH', 'MENTAL ILLNESS' AND 'WELLBEING' 204
6.6. THE CONTEMPORARY VIEW OF MENTAL HEALTH: PRAGMATIC ACCOUNTS OF LANGUAGE CHANGE 206
6.7. CONCLUSION 219

7. NAMED, LABELLED AND REFERRED TO: PEOPLE WITH MENTAL ILLNESSES IN THE MI 1984-2014 CORPUS 222

7.1. INTRODUCTION 222
7.1.1. A linguistic prescription for mental illness stigma? 226
7.2. ANALYTICAL METHOD 228
7.3. ANALYSIS 231
7.3.1. 'A person experiencing a mental illness' or 'a mentally ill person'? Exploring person-first and identity-first labels in the MI 1984-2014 corpus 231
7.3.2. 'Patients', 'sufferers', and 'victims': Exploring salient naming practices in the headline sample 244
7.3.3. 'Groups', 'cases' and 'the 1 in 4': referring to people with mental illness as quantities 255
7.4. CONCLUSION 270

8. 'SUFFERING' ILLNESSES AND 'EXPERIENCING' SYMPTOMS: WAYS OF TALKING ABOUT HAVING MENTAL ILLNESS 272

8.1. INTRODUCTION 272
8.2. METHOD FOR THIS CHAPTER 273
8.3. THE EXPERIENTIAL BASIS OF LANGUAGE AND THE LINGUISTIC BASIS OF EXPERIENCE 275
8.4. ON THE GRAMMAR OF 'SUFFERING' 277
8.5. 'EXPERIENCING' SYMPTOMS AND 'SUFFERING' FROM ILLNESSES 289

8.6. CONCLUSION 301

9. DO NEWSPAPER REPORTS ACCURATELY REPRESENT THE SYMPTOMS OF MENTAL ILLNESS? 306

9.1. INTRODUCTION 306
9.2. GROUPING MENTAL ILLNESSES 311
9.3. TRAUMA DISORDERS 314
9.4. DISSOCIATIVE DISORDERS 321
9.5. BIPOLAR DISORDER 327
9.6. OBSESSIVE COMPULSIVE DISORDERS 333
9.7. PSYCHOTIC DISORDERS 337
9.8. ANXIETY DISORDERS 343
9.9. EATING DISORDERS 347
9.10. DEPRESSIVE DISORDERS 351
9.11. EXPLORING SYMPTOMS THROUGH SKETCH THESAURUS 356
9.12. CONCLUSION 363

10. CONCLUSION 365

10.1. FULFILLMENT OF RESEARCH QUESTIONS 367
10.2. IMPLICATIONS OF THIS RESEARCH 379
10.3. LIMITATIONS 381
10.4. SUGGESTIONS FOR FUTURE RESEARCH 383
10.5. CONCLUSION 384

REFERENCES 386

LIST OF FIGURES

FIGURE 4.1. APPLICABLE LINGUISTICS AS A SYNTHESIS (TAKEN FROM MATTHEISSEN, 2012: 437) 108
FIGURE 4.2. A BREAKDOWN OF RELATIONAL PROCESSES AND THEIR PARTICIPANTS..... 125
FIGURE 5.1. LEXICAL ITEMS GROUPED BY DISORDER TYPE AS CATEGORISED IN DSM-V. 144
FIGURE 5.2. SEARCH TERM FOR THE MI 1984-2014 CORPUS..... 147
FIGURE 5.3. NEXIS SEARCH HOMEPAGE 151
FIGURE 5.4. SCREENSHOT OF A NEXIS FILE AFTER DOWNLOAD..... 153
FIGURE 5.5. SCREENSHOT OF MASTER SPREADSHEET FOR 1997..... 156
FIGURE 5.6. SUMMARY OF WORKFLOW DECISIONS FOR ILLNESS SUBCORPORA CONSTRUCTION 165
FIGURE 6.1. FREQUENCY OF 'MENTAL ILLNESS' AND 'MENTAL HEALTH' ACROSS THE YEAR SUBCORPORA 181
FIGURE 6.2. THE MENTAL HEALTH CONTINUUM AS SUGGESTED BY THE WHO DEFINITION..... 186
FIGURE 6.3. FREQUENCY OF 'MENTAL HEALTH PROBLEM/ISSUE/CONDITION' OVER THE TIME PERIOD. 189
FIGURE 6.4. FREQUENCY OF 'MENTAL HEALTH CONDITION OVER THE TIME PERIOD 190
FIGURE 6.5. FREQUENCY OF 'MENTAL HEALTH ISSUE' OVER THE TIME PERIOD..... 190
FIGURE 6.6. PERCENTAGE OF INSTANCES OF 'MENTAL HEALTH' THAT OCCUR IN THE PHRASES 'MENTAL HEALTH PROBLEM', 'MENTAL HEALTH ILLNESS' AND 'MENTAL HEALTH CONDITION' 192
FIGURE 6.7. FREQUENCY OF 'WELLBEING' OVER THE TIME PERIOD 196
FIGURE 6.8. SKETCH THESAURUS VISUALISATIONS FOR 'WELLBEING' AND 'HEALTH'..... 197
FIGURE 6.9. PERCENTAGE DIFFERENCES BETWEEN 'WELLBEING' AND 'WELL-BEING'..... 198
FIGURE 6.10. FREQUENCY OF 'WELLBEING' VS. 'WELL-BEING' 198
FIGURE 6.11. WORD SKETCH DIFFERENCE FOR 'ILLNESS' AND 'HEALTH' 200
FIGURE 6.12. WORD SKETCH DIFFERENCE FOR 'ILLNESS' AND 'PROBLEM' 201
FIGURE 6.13. WORD SKETCH DIFFERENCE FOR 'WELLBEING' AND 'HEALTH' 201
FIGURE 6.14. PROCESS OF SOCIALLY-MOTIVATED LANGUAGE CHANGE FOR 'TOILET' (TAKEN FROM TRAUGOTT & DASHER, 2002: 59) 209
FIGURE 6.15. PREDICTED PROCESS OF SOCIALLY-MOTIVATED LANGUAGE CHANGE FOR 'MENTAL ILLNESS' (BASED ON MI 1984-2014 CORPUS) DATA)..... 211
FIGURE 6.16. A PRAGMATIC ACCOUNT OF LANGUAGE CHANGE: MENTAL HEALTH TO MENTAL ILLNESS 215
FIGURE 6.17. THE CONTINUUM OF MENTAL HEALTH DISCOURSE BASED ON TERMS FROM 2013 217
FIGURE 7.1. FREQUENCY OF IDENTITY-FIRST VS. PERSON-FIRST FORMS FOR AGORAPHOBIA, BULIMIA, ANOREXIA AND SCHIZOPHRENIA)..... 235
FIGURE 7.2. FREQUENCY OF PERSON-FIRST FORMS OVER TIME IN THE MI 1984-2014 237
FIGURE 7.4. FREQUENCY OF SUFFERER, VICTIM AND PATIENT ACROSS THE YEAR SUBCORPORA IN THE MI 1984-2014 CORPUS..... 246
FIGURE 7.5. FREQUENCY OF 'SUFFERER', 'VICTIM' AND 'PATIENT' ACROSS THE ILLNESS SUBCORPORA IN THE MI 1984-2014 CORPUS..... 250
FIGURE 9.2. SKETCH THESAURUS VISUALISATIONS FOR EACH ILLNESS TYPE (CALCULATED USING SKETCH ENGINE)..... 362

LIST OF TABLES

TABLE 2.1. OVERVIEW OF EXISTING RESEARCH DISCUSSED IN SECTION 2.4 "PRESS DATA" 69
TABLE 3.1. WORDLIST FOR SAMPLE CORPUS (TOP 10 WORDS BY FREQUENCY) 81
TABLE 3.2. TOP-LEVEL SEMANTIC CATEGORIES IN THE USAS TAGSET 82
TABLE 3.3. TOP 5 SEMANTIC CATEGORIES IN SAMPLE CORPUS 82
TABLE 3.4. POSITIVE KEYWORDS IN THE MANIA CORPUS COMPARED WITH SAMPLE CORPUS (CALCULATED USING WMATRIX (RAYSON, 2009)). 86
TABLE 3.5. POSITIVE KEY SEMANTIC DOMAINS IN THE MANIA CORPUS COMPARED WITH SAMPLE CORPUS 89
TABLE 3.6. NEGATIVE KEY SEMANTIC DOMAINS IN THE MANIA CORPUS COMPARED WITH SAMPLE CORPUS 89
TABLE 3.7. TOP FIVE COLLOCATES OF 'MENTAL*' IN SAMPLE CORPUS (SPAN L4-R4) 91
TABLE 3.8. TOP CONTEXT OF 'MENTAL*' IN THE SAMPLE CORPUS 93
TABLE 3.9. TOP FIVE 4-GRAMS IN THE SAMPLE CORPUS 98
TABLE 3.10. CONCORDANCE OF THE LEXICAL ITEM 'MOODS' 100
TABLE 4.1 HALLIDAYAN METAFUNCTIONS 110
TABLE 4.2. TYPES OF PROCESSES IN A TRANSITIVITY ANALYSIS (ADAPTED FROM HALLIDAY, 1973, SIMPSON, 1993 AND JEFFRIES, 2010) 119
TABLE 4.3. TYPES OF MATERIAL PROCESS 120
TABLE 4.4. TYPES OF VERBALIZATION PROCESS 120
TABLE 4.5. TYPES OF MENTAL PROCESS 121
TABLE 4.6. TYPES OF RELATIONAL PROCESS (TAKEN FROM HALLIDAY (2004: 239)) 122
TABLE 4.7. ATTRIBUTIVE AND IDENTIFYING PROCESSES WITHIN THE RELATIONAL PROCESSES CATEGORY 123
TABLE 5.1. BREAKDOWN OF THE SEARCH TERMS. 147
TABLE 5.2. OVERVIEW OF MENTAL HEALTH LEGISLATION 1983-2013 150
TABLE 5.3. INITIAL LIST OF SUBCORPORA AND TERMS INCLUDED 161
TABLE 5.4. INITIAL WORKFLOW DECISIONS FOR ILLNESS SUBCORPORA CONSTRUCTION 162
TABLE 5.5. WORKFLOW DECISIONS FOR ILLNESS SUBCORPORA CONSTRUCTION (THIRD PHASE) 163
TABLE 5.6. OVERVIEW OF THE ILLNESS SUBCORPORA 167
TABLE 5.7. OVERVIEW OF THE YEAR SUBCORPORA 167
TABLE 5.8. CONCORDANCE PLOTS SHOWING A SAMPLE OF TERMS IN THE MENTAL ILLNESS 1984-2014 CORPUS 172
TABLE 5.9. TOP 40 KEYWORDS IN MI 1984-2014 CORPUS COMPARED WITH SIBOL ENGLISH BROADSHEET NEWSPAPERS 1993-2013. 173
TABLE 6.1. TOP 20 COLLOCATES OF 'MENTAL HEALTH' LISTED BY FREQUENCY (CALCULATED USING SKETCH ENGINE, R3-L3, MIN FREQ. = 5, MI CUT-OFF = 3) 185
TABLE 6.2. CONCORDANCE LINES FOR 'MENTAL HEALTH ISSUE' 191
TABLE 6.3. CONCORDANCE OF FOR 'MENTAL HEALTH AND WELLBEING'. 194
TABLE 6.3. CONCORDANCE FOR 'STIGMA' + 'MENTAL HEALTH' 202
TABLE 6.4. CONCORDANCE FOR 'STIGMA' + 'MENTAL ILLNESS' 203
TABLE 6.5. WORD SKETCH COLLOCATES OF "WELLBEING AND…" 203
TABLE 6.6. TOP 10 COLLOCATES OF 'MENTAL ILLNESS', 'MENTAL HEALTH' AND 'WELLBEING' (CALCULATED USING SKETCH ENGINE) 204
TABLE 7.1 EXAMPLE OF CATEGORISATION TABLE FOR NAMING PRACTICES BY REFERENT TYPE 230
TABLE 7.2. HEADLINES INCLUDING PERSON-FIRST LANGUAGE IN THE HEADLINE SAMPLE 233
TABLE 7.3. CONCORDANCES FOR 'PEOPLE WITH MENTAL ILLNESS' TAKEN FROM THE MI 1984-2014 CORPUS 239
TABLE 7.4. CONCORDANCE FOR 'PEOPLE WITH MENTAL ILLNESS' WITH CONCORDANCE LINES FEATURING THE STATISTICALLY SIGNIFICANT COLLOCATES '*MYTH*', '*DISCRIMINATION*', '*STIGMATISED*', '*BELITTLE*', '*STIGMA*' AND '*PREJUDICE*' 241
TABLE 7.5. CONCORDANCES FOR 'THE MENTALLY ILL' TAKEN FROM THE MI 1984-2014 CORPUS 242
TABLE 7.6. L1 AND R1 COLLOCATES OF 'MENTALLY ILL' (MIN FREQ. 5) IN THE MI 1984-2014 CORPUS 243
TABLE 7.7. CONCORDANCES OF 'VICTIM OF' IN THE BNC 247
TABLE 7.8. NAMING PRACTICES IN THE HEADLINE SAMPLE CONTAINING THE HEAD NOUN 'VICTIM' 248
TABLE 7.9. CONCORDANCE LINES FOR 'PATIENT' COLLOCATES 'DISCHARGED' AND 'RELEASED' IN THE SCHIZOPHRENIA CORPUS (L3, R3, MIN FREQ.5) 252
TABLE 7.10. FREQUENCY OF SALIENT NAMING STRATEGIES USED IN PLURAL AND SINGULAR FORM IN THE MI 184-2014 CORPUS 256
TABLE 7.11. REFERRING TO PEOPLE AS NUMBERS OR STATISTICS IN HEADLINE SAMPLE 257
TABLE 7.11. PEOPLE DESCRIBED AS QUANTITIES OR STATISTICS IN THE SAMPLE BY YEAR 258
TABLE 7.12. 5-GRAMS RELATING QUANTITY OR NUMBER IN THE MI-1984-2014 CORPUS 259

TABLE 7.13. COLLOCATE MODIFIERS OF 'RISE IN THE MI 1984-2014 CORPUS 260
TABLE 7.14. EXAMPLE CONCORDANCE LINES OF [MENTAL ILLNESS + RISE/INCREASE] IN THE MI 1984-2014 CORPUS. 261
TABLE 7.15. EXAMPLES OF 'CLUNIS' AND 'SILCOCK' USED TO SYMBOLIZE VIOLENCE COMMITTED BY PEOPLE WITH MENTAL ILLNESS 269
TABLE 8.1. TYPES OF PROCESSES IN A TRANSITIVITY ANALYSIS (ADAPTED FROM HALLIDAY, 1973, JEFFRIES, 2010 AND SIMPSON, 1993) 278
TABLE 8.2. EXAMPLES OF 'SUFFER' IN THE PROTOTYPICAL TEXTS SAMPLE 279
TABLE 8.3. EXAMPLE PARTICIPANTS AND PROCESSES IN MENTAL PROCESSES 280
TABLE 8.4. CONCORDANCES FOR THE PHRASE 'SUFFER FROM' TAKEN FROM THE BNC 281
TABLE 8.5. CONCORDANCE FOR THE VERB 'EXPERIENCE' IN THE BNC 283
TABLE 8.6. INSTANCES OF [FIRST-PERSON PRONOUN + SUFFER] IN THE MI 1984-2014 CORPUS 284
TABLE 8.7. INSTANCES OF [FIRST-PERSON PRONOUN + EXPERIENCE] IN THE MI 1984-2014 CORPUS 285
TABLE 8.8. CONCORDANCE FOR [I AM SUFFERING] IN THE MI 1984-2014 CORPUS 288
TABLE 8.9. CONCORDANCE ANALYSIS OF SENTENCES CONTAINING THE LEMMAS 'EXPERIENCE' AND 'SUFFER' IN THE MI 1984-2014 290
TABLE 8.10. SUBJECT COLLOCATES OF THE LEMMA 'SUFFER' IN THE MI 1984-2014 CORPUS. 291
TABLE 8.11. INVENTED SENTENCES FOR [SUFFER + PP] AND [SUFFER + NP] 295
TABLE 8.12. R1 NOUN COLLOCATES OF 'SUFFER' AND 'SUFFER FROM' IN THE BNC. 296
TABLE 8.13. FEATURE MATRIX FOR 'EXPERIENCE', 'SUFFER' AND 'SUFFER + PREPOSITION' 298
TABLE 8.14. SENSES, DEFINITION AND EXAMPLE USAGES FOR 'SUFFER' (TAKEN FROM THE OED) 298
TABLE 8.15. LEXICOGRAMMATICAL HEURISTICS FOR 'SUFFER' 300
TABLE 9.1. PROCESS OUTLINE FOR COMPARING THE SOCIAL REALITY OF THE MENTAL ILLNESS DISCOURSE DOMAIN WITH THE LINGUISTIC MANIFESTATIONS OF THE DISCOURSE DOMAIN 310
TABLE 9.2. TOP 30 KEYWORDS IN THE PTSD CORPUS COMPARED WITH ALL OTHER ILLNESS CORPORA MINUS THE 'MENTALILLNESS' CORPUS CALCULATED USING ANTCONC (ANTHONY, 2017) 317
TABLE 9.3. TOP 30 POSITIVE KEY SEMANTIC DOMAINS IN THE PTSD CORPUS COMPARED WITH THE 'MENTALILLNESS' CORPUS (LL CUT OFF 10.83, MIN. FREQ. 5, P < 0.001, *P* < 0.001) CALCULATED USING WMATRIX (RAYSON, 2008) 318
TABLE 9.4. TOP 30 KEYWORDS IN THE DID CORPUS COMPARED WITH ALL OTHER ILLNESS CORPORA MINUS THE 'MENTALILLNESS' CORPUS CALCULATED USING ANTCONC (ANTHONY, 2017) 323
TABLE 9.5. TOP 30 POSITIVE KEY SEMANTIC DOMAINS IN THE DID CORPUS COMPARED WITH THE 'MENTALILLNESS' CORPUS (LL CUT OFF 10.83, MIN. FREQ. 5, *P* < 0.001) CALCULATED USING WMATRIX (RAYSON, 2008) 324
TABLE 9.5. TOP 30 KEYWORDS IN THE BIPOLARDISORDER CORPUS COMPARED WITH ALL OTHER ILLNESS CORPORA MINUS THE 'MENTALILLNESS' CORPUS CALCULATED USING ANTCONC (ANTHONY, 2017) 329
TABLE 9.6. TOP 30 POSITIVE KEY SEMANTIC DOMAINS IN THE BIPOLARDISORDER CORPUS COMPARED WITH THE 'MENTALILLNESS' CORPUS (LL CUT OFF 10.83, MIN. FREQ. 5, *P* < 0.001) CALCULATED USING WMATRIX (RAYSON, 2008) 330
TABLE 9.7. TOP 30 KEYWORDS IN THE OCD CORPUS COMPARED WITH ALL OTHER ILLNESS CORPORA MINUS THE 'MENTALILLNESS' CORPUS CALCULATED USING ANTCONC (ANTHONY, 2017). 334
TABLE 9.8. TOP 30 POSITIVE KEY SEMANTIC DOMAINS IN THE OCD CORPUS COMPARED WITH THE 'MENTALILLNESS' CORPUS (LL CUT OFF 10.83, MIN. FREQ. 5, *P* < 0.001) CALCULATED USING WMATRIX (RAYSON, 2008) 335
TABLE 9.11. TOP 30 KEYWORDS IN THE PSYCHOSIS CORPUS COMPARED WITH ALL OTHER ILLNESS CORPORA MINUS THE 'MENTALILLNESS' CORPUS CALCULATED USING ANTCONC (ANTHONY, 2017) 339
TABLE 9.12. TOP 30 POSITIVE KEY SEMANTIC DOMAINS IN THE PSYCHOSIS CORPUS COMPARED WITH THE 'MENTALILLNESS' CORPUS (LL CUT OFF 10.83, MIN. FREQ. 5, *P* < 0.001) CALCULATED USING WMATRIX (RAYSON, 2008) 340
TABLE 9.13. TOP 30 KEYWORDS IN THE ANXIETY CORPUS COMPARED WITH ALL OTHER ILLNESS CORPORA MINUS THE 'MENTALILLNESS' CORPUS CALCULATED USING ANTCONC (ANTHONY, 2017) 344
TABLE 9.14. TOP 30 POSITIVE KEY SEMANTIC DOMAINS IN THE ANXIETY CORPUS COMPARED WITH THE 'MENTALILLNESS' CORPUS (LL CUT OFF 10.83, MIN. FREQ. 5, *P* < 0.001) CALCULATED USING WMATRIX (RAYSON, 2008) 345
TABLE 9.15. TOP 30 KEYWORDS IN THE EATINGDISORDER CORPUS COMPARED WITH ALL OTHER ILLNESS CORPORA MINUS THE 'MENTALILLNESS' CORPUS CALCULATED USING ANTCONC (ANTHONY, 2017) 348
TABLE 9.16. TOP 30 POSITIVE KEY SEMANTIC DOMAINS IN THE EATINGDISORDER CORPUS COMPARED WITH THE 'MENTALILLNESS' CORPUS (LL CUT OFF 10.83, MIN. FREQ. 5, *P* < 0.001) CALCULATED USING WMATRIX (RAYSON, 2008) 349
TABLE 9.17. TOP 30 KEYWORDS IN THE DEPRESSION CORPUS COMPARED WITH ALL OTHER ILLNESS CORPORA MINUS THE 'MENTALILLNESS' CORPUS CALCULATED USING ANTCONC (ANTHONY, 2017) 353
TABLE 9.18. TOP 30 POSITIVE KEY SEMANTIC DOMAINS IN THE DEPRESSION CORPUS COMPARED WITH THE 'MENTALILLNESS' CORPUS (LL CUT OFF 10.83, MIN. FREQ. 5, *P* < 0.001) CALCULATED USING WMATRIX (RAYSON, 2008) 354

TABLE 9.19. COLLOCATES OF QUERY TERM + AND IN THE MI 1984-2014 CORPUS (LOGDICE STATISTICAL CALCULATION, RANKED BY LOGDICE SCORE). 358

1. Introduction

In 2003, the British newspaper *The Sun* ran the front-page headline "Bonkers Bruno Locked Up". The story reported on the former professional boxer, Frank Bruno, being taken to a psychiatric hospital after being sectioned under the 1983 Mental Health Act. Bruno had been experiencing depression and was later diagnosed with bipolar disorder.

The headline generated a media furore because of the language it used to describe Bruno's hospitalisation, and because the article referred to Bruno as a 'nut'. In response to the backlash and in an attempt to right the offence caused by the headline, *The Sun* re-released the article with the new headline 'Sad Bruno in Mental Home'. Speaking to *The Guardian* newspaper about the *Sun* headline, Majorie Wallace, the then chief executive of the mental health charity *Sane*, said, "It is both an insult to Mr Bruno and damaging to the many thousands of people who endure mental illness to label him as 'bonkers' or 'a nutter' and having to be 'put in a mental home' (Gibson, 2003). After the release of the new headline, mental health professionals were still dissatisfied with the language used by *The Sun*, commenting that the use of "mental home" in reference to the psychiatric hospital Bruno was admitted to did not properly convey the purpose of such institutions (as places to recover from being ill) and said that the headline did not reflect the positive step that being sectioned can be; i.e. it is at that point that a person may finally get help (Persaud, 2003). The *Guardian* article equated the reporting of Bruno's hospitalisation by *The Sun* with a previous article the newspaper had published in 1989 reporting on the Hillsborough disaster in which 96 people died[1]. The article reported that Liverpool football fans attacked police while they tried to rescue injured fans, leading to a mass boycott of the newspaper that is

[1] The Hillsborough disaster (named after the Hillsborough stadium in Sheffield, UK) was a fatal human crush in which 96 people were killed and 766 people were injured as a result of overcrowding in the tunnels leading into the Hillsborough football ground. The event has been widely discussed ever since, with reports focusing on getting justice for the 96 people who died, after the press and the police falsely blamed the fans for the incident. As a result of the press reports, the Hillsborough disaster has become a highly contentious and even taboo subject.

still ongoing today. Discussing both the Hillsborough disaster and the representation of Bruno's hospitalisation, the journalist Owen Gibson described the negative reaction to the articles as being a result of the newspaper misjudging "the public mood" (Gibson, 2003). Gibson's comment here is an interesting one, as it suggests that there had been a shift in the public understanding or awareness of mental illness that had not been understood by the press at that time. Gibson's decision to equate the public reaction to the 'Bonkers Bruno' headline and the Hillsborough disaster demonstrates how strong Gibson felt the public reaction was to the language used in the article at that time.

Since the 'Bonkers Bruno' headline, a range of media guidelines about how to responsibly report on mental illness have been created. For example, *Time to Change*, an anti-stigma initiative launched in 2007 by the UK mental health charities *Mind* and *Rethink Mental Illness*, has a section of their website dedicated to offering advice about how to accurately portray mental illness in the media and in fictional depictions. They write:

Avoid using:

- 'a psycho' or 'a schizo'
- 'a schizophrenic' or 'a depressive'
- 'lunatic' 'nutter' 'unhinged' 'maniac' 'mad'
- 'the mentally ill', 'a person suffering from' 'a sufferer', a 'victim' or 'the afflicted'
- 'prisoners' or 'inmates' (in a psychiatric hospital)
- 'released' (from a hospital)
- 'happy pills'

Instead try:

- 'a person who has experienced psychosis' or 'a person who has schizophrenia'
- someone who 'has a diagnosis of' is 'currently experiencing' or 'is being treated for…
- 'a person with a mental health problem'
- 'mental health patients' or 'people with mental health problems'
- 'patients', 'service users' or clients

- ‘discharged’
- ‘antidepressants', 'medication' or 'prescription drugs'

(‘Mind Your Language’, *Time to Change* 2019)

Taken together, the response to the ‘Bonkers Bruno’ headline and the guidelines released by *Time to Change* demonstrate that the language used to refer to and discuss mental illness is extremely important. They also show that the public are increasingly aware of the role that language plays in public perceptions of mental illness. The introduction of media guidelines for how to write about mental illness is a positive step because it indicates that there is greater awareness of the ideological effect language can have on how the mental illness in question is perceived. However, the language that anti-stigma initiatives such as *Time to Change* prescribe is not based on any linguistic research. The lack of linguistic research into the area of prescribed forms for discussing mental illness, then, constitutes a gap in the existing research which this thesis aims to fill.

Recent reports into stigma around mental illness suggest that attitudes towards mental illness are changing for the better. For example, Corker et al. (2016) report that discrimination “has fallen significantly” (2016: 6) in recent years, which they suggest may be related to (but cannot be directly attributed to) the *Time to Change* anti-stigma campaign. The research conducted by Corker et al. (2016), like the majority of previous research into the stigma surrounding mental illness, is based on analyses that make use of methodological tools such as bespoke scales (Corker et al. 2016 use the Discrimination and Stigma Scale). Such gradations are often based on Likert-type scales; e.g. a statement alongside a four-point scale that the participant can agree or disagree with, using the scale to denote the degree of their agreement/disagreement). Nonetheless, such studies are typically interested in attitudes towards mental illness, and not in how language affects these attitudes.

In addition to the research that uses scales to assess attitudes to mental illness, much previous research has also used bespoke coding schemes for analysing

stigmatising themes in newspaper discourse. For example, Rhydderch et al. (2016) explored the effect that the *Time to Change* campaign had between 2008-2014 using a coding scheme to inform a content analysis (i.e. a process by which a text is analysed for its constituent themes). After analysing the results using univariate logistic regression models, Rhydderch et al. (2016) concluded that there was a decrease in stigmatising articles, with "an increase in the proportion of antistigmatising articles which approached significance at $p < 0.05$" (Rhydderch et al., 2016:5). The analysis conducted by Rhydderch et al. (2016) (i.e. a content analysis with the aim of finding stigmatising articles which is then quantified for statistical analysis) is typical of the existing research into press representations of mental illness, in that (i) the focus is on locating stigmatising articles, (ii) the method is thematic in the first instance and then quantitative in the second, and (iii) there is no or very little exploration of language (Rhydderch et al. 2016 do hypothesise that there will be a decrease in 'pejorative language' but they do not explain what they mean by this term).

The widespread use of scales and coding models (which I discuss in more detail in Chapter 2) means that the analysis of representations of mental illness in the press is typically quantitative and does not privilege language as an object of study. This has led to an extraordinary situation in which there exist prescribed linguistic forms for discussing mental illness (i.e. those listed above by *Time to Change*) and very many studies on the stigma surrounding mental illness in the press, yet no research into how these things are related; i.e. how prescribed linguistic forms may affect stigmatising attitudes. A lack of research also means that it is unclear whether there really is a linguistic basis for these prescriptions. There is an implicit assumption in the existing research that language does indeed affect how people view mental illness (otherwise why prescribe language for talking about mental illness and why generate hypotheses based on language use?) and yet there exists no comprehensive account of how, linguistically, mental illness is represented the press. This is the context which has led to this thesis.

Despite the fact that previous research demonstrates an improvement in attitudes towards mental illness, it is now 16 years since *The Sun*'s headline about Bruno and it is still the case that media depictions are often inaccurate, over-simplistic and stigmatising. A recent example of a damaging media representation from February 2019 concerns the death of the musician Keith Flint. Against extant advice developed by *The Samaritans* on how to responsibly report celebrity suicide in the media, *The Daily Mail* reported Flint's death by giving specific details of the method of suicide and *The Sun* reported the story via a front-page headline. Both newspapers stated that the cause of Flint's suicide was depression caused by the breakdown of his marriage. *The Samaritans* advise that celebrity suicides should not be sensationalised, reported as a front-page item or be attributed to simplistic causes. Moreover, research conducted by *The Samaritans* has shown there to be a link between media coverage and suicide rates with suicide rates increasing significantly if "suicide methods are reported, if the story is placed prominently and if the coverage is extensive or sensationalised" (*The Samaritans*, 2019). The language used to describe mental illness, then, has real-world and potentially fatal consequences. Moreover, mental illness is increasingly prevalent in the UK. Statistics show that ¼ of people in the UK will experience a mental health problem each year, with 1 in 6 people in the UK reporting a mental health problem each week[2] (*Mind*, 2019a). In addition to this, because of the existing stigma of mental illness (and the resultant self-stigma) caused by the language surrounding mental illness, the rate at which people experiencing mental health problems access and continue to access healthcare is hindered (Schomerus et al., 2012; Flynn et al., 2016). It is also the case that, despite a tendency to link mental illness with crime and violence in the press, 45% of people with a mental illness will be a victim of crime themselves, with people with schizophrenia being more likely to be victims of

[2] The statistics taken from Mind (https://www.mind.org.uk/information-support/types-of-mental-health-problems/statistics-and-facts-about-mental-health/how-common-are-mental-health-problems/#two) are based on people living in private housing in England. As a result, they may not be indicative of the mental health problems faced by "people in hospital, prison, sheltered housing, or people the homeless" (*Mind*, 2019a).

a crime than perpetrators of it. For the reasons outlined in this section, there is much scope for research into the language of mental illness in the press and the potential ideological effects that language choices can have. This is the topic of this thesis. In the next section I briefly summarise the value and importance of studying media discourse from a linguistic perspective to gain a perspective on the language of mental illness. Following this, I outline the research questions (RQs) that my thesis answers.

1.1 The language of mental illness

Fairclough (1995) states that "media texts constitute a sensitive barometer of sociocultural change, and they should be seen as valuable material for researching change" (1995: 52). Newspaper discourse, then, provides insight into the language of mental illness for reasons wider than those detailed in the previous section. Studying media discourse enables us to look not just at the reporting of individual events (such as the case of Keith Flint) but also at the linguistic patterns that are indicative of public perceptions of mental illness, such as the labels used to describe specific mental illnesses and the people that 'have' mental illness. For example, there is a significant difference between describing someone as 'a person with schizophrenia' and 'a schizophrenic'; similarly, there is a difference between calling someone 'a patient' and a 'sufferer' of mental illness (I explore these two issues in more detail in Chapters 7 and 8 respectively). Moreover, studying media discourse diachronically allows us to chart how these usages and their effects have changed over time. Through linguistic analysis, it is possible to study empirically the 'public mood' that Gibson (2003) invoked to explain the outrage at the 'Bonkers Bruno' headline.

In recent years a wealth of research has been conducted into the language of mental health in the press. The vast majority of this research, however, has been carried out by researchers working in the fields of psychiatry or media studies and, consequently, has not deployed the systematic tools for language analysis developed

in linguistics. Moreover, the few analyses that have been conducted by linguists have tended to be qualitative or specific to one theme in mental health research, resulting in small-scale studies that are open to criticism, particularly the over-reliance on the subjective interpretation of findings. Generally, previous studies of how mental illness is discussed in the press tend to focus on a particular theme (e.g. analysing stigma) or on a particular mental illness (e.g. schizophrenia). These studies tend either to find evidence that stigma exists (although this is perhaps unsurprising given that studies analysing stigma already presuppose its existence) or only offer valuable information about a specific mental illness. In contrast, in this thesis a range of mental illnesses (as defined by the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition, hereafter DSM-V) are analysed with no *a priori* expectations of their representation. For example, while my research questions are not concerned specifically with stigma, they do not neglect the possibility (and in fact high probability) that stigmatising language may be found. It is my opinion that corpus-assisted discourse analysis that asks specific questions of the data in question (e.g. 'how is stigma manifest in language?') only succeeds in corroborating the existence of top down social constructs since it does not explore the data from the bottom-up, which is what is needed in order to explore how meaning is constructed in the data. Furthermore, the analyst already knows that such constructs are present in the data because the varied nature of a specialised corpus requires that the analyst collect data using specific search terms. As a result, the analyst already knows that certain terms related to a particular construct, e.g. stigma, will be present in the data. To remedy criticisms of such interpretative positivism in corpus analysis, it is necessary to broaden research questions (and data collection) significantly more widely than previous studies have. Doing this offers a means of remedying the claims of positivism inherent in corpus linguistic analysis (i.e. the notion that a corpus can only tell you what is in it and not what is absent), as well as advancing our understanding of mental health and illness reportage. The reason for this is that the topics discussed in articles concerning mental illness are

significantly more diverse than stigmatising language; for example, newspaper reports of mental illness and arts initiatives (Atanasova et al. 2019)[3].

1.2 Research questions

Having outlined the context for this study, and the importance of studying the linguistic representation of mental illness in news reports, I now present the research questions that I answer in this thesis. These are as follows:

1. How are the terms 'mental illness' and 'mental health' used in the MI 1984-2014 corpus?
2. What linguistic strategies are used to name, label and describe people with mental illness in the MI 1984-2014 corpus?
   2.1. To what extent is person-first language present in the MI 1984-2014 corpus?
   2.2. What themes are present in the corpus for referring to people with mental illness?
3. What processes are associated with mental illness in the MI 1984-2014 corpus?
   3.1. What terms do the press use to refer to having mental illness?
4. Is the depiction of mental illness realistic?
   4.1. Are the symptoms of each disorder type (e.g. depressive illnesses) accurately portrayed in the press?

Answering these questions requires both qualitative and quantitative analysis, as well as top-down and bottom-up approaches. The analysis of the basic linguistic units of the noun phrase and the verb phrase in RQs 2 and 3 constitutes a bottom-up analysis because I am not starting my analysis by looking any *a priori* feature above what

[3] Despite this, research to date has focused on negative press representations.

naming strategies or processes are frequent in the corpus. By contrast, RQs 1 and 4 are more top-down because I start my analysis with the intention of exploring a predetermined feature of the texts, i.e. how 'mental health' and 'mental illness' are used in the corpus (RQ1), and whether the symptoms of illness are present in the corpus (RQ4).

Answering these research questions also requires the use of a representative corpus. In the next section I introduce the MI 1984-2014 corpus and show how I designed the corpus to allow me to answer the questions listed in this section.

1.3 The Mental Illness Corpus 1984-2014

I discuss the corpus creation procedure in detail in Chapter 4; however, I will briefly introduce the data here. The corpus comprises 50,729,32 words of newspaper articles reporting on mental illnesses between 1984 and early 2014. The timeframe of the corpus saw a relatively high level of mental health reform. For example, The Mental Health Act 1983, The Mental Health (Patients in the Community) Act (1995), The Mental Capacity Act (2005), the amendments to the 1983 Mental Health Act in 2007 and the 2013 Mental Health (Discrimination) Act all occurred within the time period covered by the corpus. The dates covered by the MI 1984-2014 corpus, then, are broad enough to provide insight into any diachronic change in the lexis and structural forms associated with mental illness, as well as being fitted to a relevant period in UK mental health history. The MI corpus is, then, continuous and longitudinal.

The articles were collected using the LexisNexis database. The search terms were collated using information from the *Mind* 'A-Z of mental illnesses' (*Mind*, 2018) and informed by the 5th Edition of the *Diagnostic and Statistical Manual of Mental Disorders* (2013) (DSM-V). To allow for maximum coverage of mental illnesses and their associated lexical forms, i.e. nominal forms such as *psychosis* and adjectival forms such as *psychotic*, wildcards (such as !) were used; e.g. *psycho!*. After construction of

the search terms, the following sampling frame was used to search the LexisNexis UK Newspapers database from January 1983 to January 2014[4] (articles were grouped for high similarity):

> (mental illness!) OR (mental health!) OR (mental ill health) OR (mentally ill) OR (mentally un!) OR (agoraphobi!) OR (anorexi!) OR (anxiety) OR (autism) OR (autistic) OR (binge eating disorder) OR (bipolar!) OR (body dismorph!) or (borderline personalit!) OR (BPD) OR (bulimi!) OR (depress!) OR (dissociative disorder) OR (dissociative identity disorder) OR (eating disorder) OR (multiple personality disorder) OR (mpd) OR (obsessive compulsive disorder!) OR (obsessive compulsive) OR (ocd) OR (paranoia) OR (personality disorder) OR (postnatal depression) OR (posttraumatic stress) OR (post traumatic stress) OR (post-traumatic stress) OR (ptsd) OR (psychosis) OR (psychotic) OR (schizophreni!) OR (seasonal affective disorder!) OR (social phobia) OR (bulimia) OR (hypomania) OR (hypermania) OR (mania) OR (mania!) OR (manic) OR (manic!) OR (schizo)

The corpus creation procedure facilitated the creation of illness subcorpora in addition to year subcorpora, guided by the original search terms; for example, the terms collected as part of the eating disorder and OCD subcorpora are shown below.

Eating Disorder subcorpus	eating disorder, eating disorders*, bulimi*, binge eating disorder, anorexi*
OCD subcorpus	obsessive compulsive disorder*, obsessive compulsive, ocd

The design of the corpus and the illness subcorpora allowed for the analysis of diachronic (across time) as well as synchronic (e.g. across illness) variation.

The Mental Illness Corpus 1984-2014 that I created to answer my research questions constitutes a significant resource for the investigation of constructions of mental illness in the UK press. In the next section, I explain how it is used in each of the subsequent chapters of this thesis.

[4] No data was collected for 1983 as the database returned no hits for this time period.

1.4 Structure of this thesis

In the next Chapter (Chapter 2) I provide a review of the existing literature on the representation of, and attitudes towards, mental illness in a variety of text types, e.g. online data, newspaper data and spoken data. In addition to research on the representation of mental illness in these different data types, I also discuss the theoretical position of Social Constructionism (particularly in reference to CDA). My review of the literature in Chapter 2 includes coverage of research from psychiatry, linguistics and counselling. As a result of including the research conducted into a wide variety of different text types, in addition to a wide variety of analytical fields (e.g. psychiatry, linguistics), the literature review in Chapter 2 is very broad. The reason I have included such a broad variety of literature is that no study to date has surveyed this range of work for what it says about the linguistic representation of mental illness. Moreover, due to the fact that the data contained in the MI 1984-2014 corpus is so broad (in that it covers a significant period of time as well as a significant number of illnesses), I felt it important to thoroughly review the existing literature into a range of illness types as well as analytical methods.

In Chapter 3, 'Analytical Methods 1: Corpus Linguistics', I provide a brief overview of corpus linguistics with particular reference to the approach I adopt in this thesis: namely, corpus linguistics as a method (as opposed to corpus linguistics as a theory) (McEnery & Hardy, 2012; Tognini-Bonelli, 2001). I also detail the specific corpus linguistic analytical methods that I use in this thesis, such as collocation and keyness analysis and the statistical tests and cut-offs associated with each analysis. Due to the fact that I exemplify each analytical method using data from the MI 1984-2014 corpus (specifically the data collected during a pilot study and an illness-specific sample of the data), I also illustrate the utility of each analysis type for analysing ideology in texts. Chapter 3, then, offers an overview of the corpus tools I use in this thesis but also provides information about the theory underpinning each analysis. For

example, a node word's collocates (found by conducting a collocation analysis) provides evidence about the meaning of that node word, as in Firth's statement that "you shall know a word by the company it keeps" (Firth, 1957: 11).

Following a similar structure to Chapter 3, in Chapter 4, 'Analytical Methods 2: Critical Discourse Analysis', I provide an overview of Critical Discourse Analysis, covering the early manifestations of linguistic inquiry into ideology in texts such as that of the East Anglia School (Fowler et al., 1979) to contemporary research into corpus-assisted discourse analysis that combines these early principles of CDA with computational methods. I also discuss and contest the notion that the automation of textual analysis offered by corpus linguistics provides a magic bullet for objectivity in CDA. As was the case in Chapter 3, I also outline the different CDA methods that I use in this thesis in Chapter 4. Specifically, I discuss, and exemplify using relevant data, Halliday's transitivity model, taken from his model of Systemic Functional Linguistics (1973), and naming analysis.

In Chapter 5, 'Corpus Construction', I describe the process of constructing the MI 1984-2014 corpus, from compiling the sampling frame (e.g. search terms, dates covered) to the compilation procedure for the illness and year subcorpora. In particular, I provide a detailed discussion of the interpretative status of search terms and show how the dates the corpus covers are fitted and relevant to the topic matter in terms of legislation passed in this period. In my report of the corpus construction procedure, I cover practical issues related to compiling corpora such as cleaning the data. Furthermore, I include a discussion of the problems that the interdisciplinary nature of corpus construction poses for the researcher (i.e. that the researcher has to know about programming, statistical methods, etc. in addition to methods of linguistic analysis). I conclude Chapter 5 by showing how the corpus construction procedure resulted in a representative and evenly distributed corpus.

Chapter 6 is the first analysis chapter. In Chapter 6, I answer RQ1, "Has the meaning of mental illness changed over time in the press?". I start my analysis in Chapter 6 by arguing that the terms 'mental health' and 'mental illness' have been

used interchangeably in the previous literature on the representation of mental illness in the press. Specifically, I argue that using these two terms interchangeably (especially during data collection) may result in incomparable datasets. In the rest of Chapter 6, I argue that the terms 'mental illness' and 'mental health' are distinct terms, and that the meaning of the two terms has shifted over the time period covered by the MI 1984-2014 corpus. I argue that the lexical change I observed is consistent with pragmatic accounts of language change in which the language development is in part a result of euphemism (e.g. Traugott & Dasher, 2002).

In Chapter 7, I explore the labels associated with mental illness in more detail, specifically through naming analysis. In doing this, I address RQs 2-2.2 which are concerned with how people with mental illness are named and what themes are present related to naming practices in the MI 1984-2014 corpus. In Chapter 7, I discuss prescribed forms for referring to people with mental illness (such as person-first language) and explore the frequency of such prescribed forms in the corpus. In addition. I explore salient naming strategies in the corpus, particularly the labels 'patient', 'sufferer' and 'victim'. Using corpus evidence, I show that these labels are patterned to specific illness types. Furthermore, I examine themes in naming practices in the MI 1984-2014 with particular reference to referring to people with mental illness as numbers. I argue that the tendency in the corpus to refer to people as quantities and statistics depersonalises people with mental illness. Moreover, I argue that the 'rhetoric of quantification' (Fowler, 1991: 166) provides a way for the press to sensationalise news events related to mental illness which in turn constitutes the representation of mental illness as a 'moral panic' (Cohen, 1973).

In Chapter 8, I address RQs 3-3.1 to explore the ways in which the press talk about people having mental illness. In order to do this, I conduct a mixed-methods analysis of the data. In the chapter, I analyse the salient processes described in a sample of the data and then explore the frequency and the semantic and pragmatic content of the verbs 'suffer' and ' experience' in the context of prescribed forms for talking about having mental illness (e.g. the *Time to Change* advice to avoid the verb

'suffer'). In Chapter 8, I show that 'suffer' and 'experience' occur in different semantic contexts in the MI 1984-2014 corpus as well as general language corpora, which may contribute to 'suffer' being a more problematic term for describing mental health than 'experience'. Moreover, I show that 'suffer' is proportionally less likely to be used in first-person narratives because 'suffering' is attributed to people with mental illness by others, e.g. medical professionals, in reported speech. I bring together my findings in a set of lexicogrammatical heuristics based on the semantic content of 'suffer' and 'experience' in context (e.g. whether the word encodes animacy or is temporally bounded). I argue that my analysis provides a more nuanced understanding of what 'suffer' in its transitive and intransitive form means, both in terms of who and what it can be used in reference to.

In a departure from the bottom-up analyses conducted in Chapters 6-8 (i.e. analyses that are motivated by a particular linguistic feature, such as the noun phrase or the verb phrase), in Chapter 9 I investigate if and how the symptoms of mental illness are present in the MI 1984-2014 corpus. In doing this I address RQ4 and RQ4.1: "Is the depiction of mental illness realistic?" and "Are the symptoms of each disorder type (e.g. depressive illnesses) accurately portrayed in the press?". In order to answer these research questions, I explore the symptoms of each disorder type covered by the corpus. Specifically, I use keyword and key semantic domain analysis to explore whether the symptoms of mental illnesses are accurately represented in news articles on mental illness. In addition to corpus tools, I also qualitatively analyse the most prototypical text for each illness subcorpus (i.e. the text that contains the most frequent features of the illness subcorpus overall) to explore whether the keyness findings are also a feature of whole texts. In Chapter 9 I also explore whether the language used in the illness subcorpora relating to illness with the same or related symptoms features lexis that overlaps.

In Chapter 10, 'Conclusion', I provide a discussion of the findings reported in this thesis and revisit my research questions. I detail the caveats of my research as well

as provide ideas for future research in the area of the language of mental illness. I also discuss the practical implications of my work.

1.5 Conventions

Throughout this thesis I make use of the following typographical conventions:

- 'single quotes' to refer to lexical forms (e.g. 'mental illness')
- "double quotes" to refer to extracts from the data, extracts from newspaper articles, or verbatim extracts from this thesis or the existing literature
- *Italics* to refer to concepts, publication names (e.g. *The Sun*), the names of organisations (e.g. *Mind*) and for emphasis, e.g. no lexical item is *inherently* stigmatising"
- SMALL CAPS to refer to semantic domains (e.g. 'flute' and 'piano' are lexical items associated with the semantic domain of MUSIC)

1.6 Conclusion

In this introduction I have shown that mental health and illness is an increasingly important topic in UK society, both in terms of the number of newspaper articles covering mental illness-related issues, and the increased prevalence of mental illness generally. I have also shown how the public are increasingly aware of the language used to discuss mental illness in the press (see, for example, the "Bonkers Bruno" and Keith Flint examples in Section 1.1). Moreover, I have explained how the language used to discuss mental illness is being increasingly prescribed by anti-stigma initiatives. Despite all of these activities and initiatives, very little research exists that

explores the language used to discuss mental illness in the press using a purely linguistic approach. For this reason, I have set out the research gap in the existing literature that this thesis goes some way to addressing. In this chapter I have also introduced the MI 1984-2014 corpus and provided an outline of what I will discuss in the rest of this thesis.

In the next chapter I review the existing literature relevant to this thesis as means of providing a detailed context for the analysis that follows in subsequent chapters.

2. The language of mental illness

2.1 Introduction

In recent years, increasing numbers of researchers in linguistics, psychiatry and mental health studies more broadly have turned their attention to the analysis of the ideological effect language has on how the public view mental illness. Prior to this, research into language and mental illness broadly fell under the umbrella of interactionalist research that makes use of spoken data (e.g. Antaki, 2007; McCabe et al., 2002). Mental illnesses are defined by the American Psychiatric Association (hereafter APA), as "health conditions involving changes in emotion, thinking or behaviour (or a combination of these). Mental illnesses are associated with distress and/or problems functioning in social, work or family activities."(APA, 2019). In this chapter, I offer an overview of research into the language of and about mental illness and position my own work in relation to it.

The following sections provide an overview of the main themes in existing literature, starting with social constructionism (Berger & Luckmann, 1966), the theoretical underpinning that most applied studies make use of (albeit without stating this fact). Given that mental illness research necessarily has to span a wide range of disciplines, studies will be reported from a range of areas, although most research is published in journals that list psychiatry as a key area. The reason for including a wide range of sub-disciplines and study types in my review of the relevant literature is to give a well-rounded perspective of what research has been conducted in the area of language and mental illness. Another practical reason for doing this is that many of the studies combine methods from various disciplines; e.g. linguistic studies use methods from sociology, and psychiatric studies adopt methods from sociology and linguistics.

In Section 2.2, I briefly outline social constructionism and how it has been used in studies relating to the analysis of language. In Sections 2.3 and 2.4, I provide an overview of the existing research into language and mental illness.

Overall, in this chapter I show that there has been much research in the area of mental illness, particularly in reference to stigma and discrimination. However, I argue that existing research into the representation of mental illness in the press neglects to identify language as an object worthy of study in itself, or as a factor in how stigma and discrimination are manifested. This chapter then, advances my thesis by identifying the gaps in current research that I argue linguistic analysis of the kind conducted in this thesis can go some way to filling. Moreover, this chapter provides a broad overview of previous research findings including findings that may not appear to be directly related to news reports on mental illness, but which I argue help to contextualise public attitudes towards mental illness. This chapter, then, also constitutes a reference point for existing research findings that I refer back to at various points in the thesis.

2.2 Social constructionism

In linguistics, much research relies on the ideas underpinning social constructionism (it is social constructionism that forms the foundation of the whole notion of 'discourse' that underpins CDA) and yet very few researchers make this link clear. Burr (1995) makes this point, stating that there has been

> a gradual emergence of a number of alternative approaches to the study of human beings as social animals. These approaches have appeared under a variety of rubrics, such as 'critical psychology', 'discursive psychology', 'discourse analysis', 'deconstruction' and 'poststructuralism'. What any of these approaches have in common, however, is what is now often referred to as social constructionism.
>
> (Burr, 1995: 1)

The tendency for researchers not to report that they subscribe to social constructionism is perhaps due to the fact that the link between language and the nature of meaning is so entrenched that to make the link explicit is unnecessary. However, it is important to detail what social constructionism is and why it is the primary theory used in studies analysing mental illness and language because subscribing to social constructionism implies a wide range of theoretical assumptions that are otherwise left unsaid. For example, the very definition of what constitutes a social construct (e.g. a text), or the belief that through the study of a social construct we have access to a society's understanding of social constructs, is often left unsaid in previous research.[5]

In its most basic form, social constructionism is concerned with the nature of knowledge, and posits that reality is constructed through our interaction with social objects and social actors. Burr (1995) states that in order to be a social constructionist one would have to believe that (i) a critical stance is needed in order to better understand "taken-for-granted ways of understanding the world" (Burr, 1995:2), (ii) ways of seeing the world are "historically and culturally relative" (1995: 2) with products being "specific to particular cultures and periods of history" (1995: 2). (iii)

[5] The idea that social constructs reflect society is one that is taken for granted in corpus linguistics, where media texts are often presented as social constructs and as being indicative of a societal view on a topic, usually without any theoretical underpinnings to these assumptions being offered. I am not raising this point to cast doubt on whether this is a sound basis for analysis but one should consider these assumptions, particularly in corpus linguistic studies, where in most cases the focus is almost entirely on the production of texts rather than the reception of them.

knowledge is co-constructed, i.e. meaning making is inherently social and interaction based, and (iii) "knowledge and social action are interlinked" (1995: 5) [6] with constructions being bound by what is permissible within a particular social setting, e.g. one may have a different understanding of a particular social construction but the possible social actions associated with that construct are limited by what is permissible within a given social setting. To exemplify this last point, Burr (1995) uses the example of societal views on drunkenness changing from seeing it as a crime to considering it an illness. In the former case, imprisonment was the acceptable social action; in the latter, the acceptable social action is medical or psychological intervention (Burr, 1995: 5).

The basis of the social constructionist view (and many other philosophical underpinnings of linguistics more generally) was greatly influenced by the work of the American Pragmatists such as Mead (1934) and Peirce (1906). Their work on the area of symbolic interactionism sees meanings as "social products, as creations that are formed in and through the defining activities of people as they interact" (Blumer, 1969: 5), as opposed to structuralist theories of meaning, which assume that something has an inherent meaning. Due to the fact that social constructionism states that reality is socially constructed, it is a useful paradigm for challenging normative or common sense understandings of society. Social constructionism, then, is concerned most primarily with the ways in which "the world can be socially constructed by the social practices of people but at the same time experienced by them as if the nature of their world is pre-given and fixed" (Burr, 1995: 13)[7].

Social constructionism has been adopted in a wide range of linguistic sub-disciplines such as sociolinguistics (Britain, 2018; Coulmas, 2016; Wolfram, 2015), linguistic ethnography (Bucholtz & Hall, 2005; Ochs & Shohet, 2006; Rampton &

[6] It is easy to see here the clear links with social based theories of language and pragmatics more generally.

[7] As Burr points out, social constructionism itself is now a social construct, and by writing a book on it he is "contributing to what might be called 'the social constructionism of social constructionism'" (Burr, 1995: 13).

Charalambous, 2016) and pragmatics (File, 2018; Marra, 2012) in varying degrees of philosophical depth. However, in this chapter, I will only offer an overview of research using social constructionism in relation to studies in CDA, and more specifically in mental illness research. My reason for having detailed above where social constructionist theory is used in linguistics more broadly is to show that the theory is well-used as a basis for language analysis, and that it does not belong to a particular branch of linguistics or to research that has particular aims (e.g. to be critical), as some critics have suggested (Hacking, 1999).

2.3 Social constructionism and CDA

Despite being the theory underpinning much work in CDA, very few CDA researchers explicitly state that their research is underpinned by a theory of social constructionism. For example Fairclough (1992) accepts that his viewpoint is one of social constructionism (or constructivism as he terms it), but attributes much of the theory underpinning his work to the constructionist thinking of Foucault (1972) (Fairclough, 1992: 37-61). Indeed, it is to Foucault that the term *discourse* is generally attributed, which is the term for language-in-use that is used frequently in linguistic social constructionist research. In contrast to the social constructionism used in discourse analysis generally, which follows Burr (1995) and Berger & Luckmann's (1966) notion of social constructionism (i.e. the analysis of how social constructs are created and negotiated through interaction and therefore how we understand 'reality' through social constructs), Foucault's conception of social constructionism is focussed almost entirely on how discourses frame "distinctive disciplinary formations through which power/knowledge (power as knowledge/knowledge as power) operates" (Clarke, 2005:149). For Foucault, the function of discourses is to produce, legitimate and maintain power asymmetries. Foucault's concern is not with the analysis of individual instances of how discourses are created *per se*. As a result, we can think of Foucault's conception of discourse as being at the macro end of the CDA spectrum,

the end that deals with sociological phenomena. In contrast, the linguistic analysis of discourses can be conceptualised as being at the micro end. The extent to which one attributes social construct status on this scale can be thought of as being relative to the analyst's frame of reference. What I mean by this is that, for Foucault, a discourse (i.e. a collection of ideas and social actions about a topic) may constitute a social construct. By comparison, for the linguist a social construct may be much less widespread and, perhaps as a result of this, more measurable, e.g. the use of particular linguistic variables to index a particular group identity. In fact, Burr (1995) makes a similar distinction between macro social constructionism or Foucauldian discourse analysis and micro social constructionism or discursive psychology (1995: 21). By drawing a distinction between micro and macro discourse analysis, I do not wish to attribute a higher status to either form. I see both as types of discourse analysis, and I do not think they are mutually exclusive. It is the case however, that I as a linguist am more interested in the micro end of the scale.

Moreover, as a result of how broad it is, Foucault's theory of discourse alone is not sufficient to provide the parameters for linguistic analysis; rather it presumes the existence of a discourse. As a result of this, Fairclough suggests that his own view of CDA involves operationalizing Foucault's theory of 'discourse' in a loose sense[8]. In later work, Fairclough claims to subscribe to a social realist paradigm because although social products are constructed, once constructed they become reality (Fairclough, 2003: 8; Sayer, 2000). Criticisms of social constructionism as anti-realist are common (i.e. that social constructionism does not account for the fact that a real social world "exists independently of our knowledge about it" (Fairclough, 2010: 355)); however, my own view on this is that a weaker version of social constructionism that takes account of brute facts (those facts that exist outside of human interaction)

[8] Foucault's conception of discourse is a sociological theory, and as such is abstract and not linked to particular social constructs, e.g. linguistic artifacts. This is a source of criticism for Fairclough. Furthermore, for Foucault, discourse is mainly concerned with power imbalance and the description of sociohistorical 'discourse formations' (Fairclough, 1992; Foucault, 1972).

and institutional facts (those facts that exist as a result of human interaction) can account for such criticisms. An oft-cited example to explain brute and institutional facts is the concept of cash money. Take for example a £10 note. The brute facts about the £10 note are that it is made of paper and contains ink. The institutional facts about the note are that it is institutionally valuable to a particular sum and we can buy things with it based on this fact. In order for the note to be more than paper and ink, we have to observe the institutional facts about the note which include (among others) financial institutions and local and global rules of commerce. In line with Searle (1995), I suggest that once a social construct is established it can take on epistemologically objective status whilst also being ontologically subjective (i.e. a social construct takes on the guise of reality through entrenchment but it is only afforded that guise by the very nature of knowledge being socially constructed).

Studies that take a social constructionist perspective are prevalent in research into mental illness. The reason for this is that such a theory provides a link between texts (i.e. newspaper articles), social actions (i.e. policy change) and the nature of knowledge (public understanding of mental illness), but also shows how mental illness as a social construct itself is diagnosed. For example, Georgaca (2013) provides an overview of how social constructionist research has provided critiques of psychiatric diagnosis. She writes that the social constructionist paradigm "attempts to denaturalize phenomena that have come to acquire a taken-for-granted character by highlighting the processes through which these are socially constituted." (2013: 56). Consequently, research into mental illness that takes a social constructionist perspective

> focuses on highlighting the contingent, socially produced character of categories of mental distress and of associated professional practices. Within this paradigm, thus, classification, the dominant system of knowledge regarding mental distress, and diagnosis, the practice of assigning a psychopathological category to a person, are not taken as given or as resources, but rather are treated as topics of investigation in their own right. The aim is to

> examine how these systems of knowledge and practice have come to take their current form, how they are accomplished in practice and finally the consequences for mental health institutions and for individuals in distress.
>
> (Georgaca, 2013: 56).

Georgaca argues that by challenging underlying common-sense understandings of mental illness diagnoses, stakeholders in mental illness activism can provide alternative, more empowering understandings of mental illness. Further to this, Conrad & Barker (2010) state that social constructionist research in the field of medical sociology has made "significant contributions to our understanding of the social dimensions of illness" (Conrad & Barker, 2010: 567). However, the use of social constructionist theory in medical settings is not without its problems; as Bury (1986) states, if we were to take a purely social constructionist approach to medicine "the stable realities of the human body and disease are in fact 'fabrications', or 'inventions' rather than discoveries" (Bury, 1986:165). Bury (1986) raises an interesting point here about the intersection between the objective nature of medicine and the constructed nature of knowledge; however, I argue that taking a social constructionist perspective, particularly in the analysis of language about mental illness is useful because, for instance, it allows for the comparison of the effects of changing labels for mental illness, e.g. prescribed language for referring to people with mental illness, and it allows for the culturally-specific aspects of mental illness to be uncovered.

For Bury (1986), the issue with constructionist approaches to medical sociology is that they negate the fact that there are some "stable realities" in medicine. Whilst I agree that there are certainly "stable realities" in medicine, perception plays a huge role in how we view illness. There is a wealth of research that demonstrates cultural differences in how people view illnesses. For example, Furnham & Baguma (1999: 121) showed that Ugandan students believed supernatural powers influenced their health to a certain extent whereas students from Great Britain did not. Sayakhot et al. (2012) found that Australian women associated the menopause with aging whereas Laotian

women did not. Gray et al. (2009) found that compared with Chinese carers of people with Alzheimer's Disease, Latino carers attributed the disease to the person having had a hard life (Gray et al., 2009: 2) and Cohen (1995) found that in India dementia was called "hot brain" (Cohen, 1995: 314) and was associated with anger, rather than memory loss. Moreover, research has shown that not only do associations with illnesses vary across cultures, but so too does the notion of what health is, which means that what a person sees as a symptom of an illness may vary cross-culturally (Wallin & Ahlström, 2010). This is hugely important because without a set of symptoms then there is no 'objective' disease, if we are to accept Bury's (1986) point that medicine should be based on "stable realities".

Moreover, as Conrad & Barker, 2010: S70) point out, some diseases, such as fibromyalgia and chronic fatigue syndrome, are contested as they are 'medically invisible' and therefore much of our understanding of them comes from the person experiencing symptoms and not medical professionals. Social constructionist approaches take account of the cultural understandings of illnesses because they take into account the social meaning of the *illness* as distinct from the biological condition or *disease*. (Conrad & Barker, 2010; Eisenberg, 1977). This is not to say that the disease side of the illness/disease distinction cannot be accounted for in a social constructionist paradigm. Conrad & Barker (2010: 68) argue that "the *disease* side of the disease/illness conceptual distinction is also ripe for social constructionist analysis, insofar as what gets labeled a disease or qualifies as biological is often socially negotiated."

Furthermore, social constructionism can account for stigma around illnesses and the concept of disability where it is often the case that people do not identify as disabled but rather see society as disabling them (Scope, 2019). The politicisation of disability is just one place that we see the social construction of disability. For example, recent changes to Income and Support Allowance in the UK requires people to have their *level* of disability assessed. Social constructionism, then, is a widely used and

useful theory to analyse the language of illness, and in the case of this thesis, the language of mental illness specifically.

In the next section, I report the existing research into mental illness and language.

2.4 Themes in research on mental health and language

In section 2.2, I described social constructionism as the theoretical basis for many of the studies into the language of mental illness. In the following sections I outline the existing research into mental illness and language. I do so in broad themes, e.g. research into stigma, research using press data. Due to the fact that I present the existing literature by theme, the different issue each subsection covers may not be on the same conceptual level; e.g. stigma is on a different conceptual level to press data. The reason I include research on stigma as a separate section is because much research in this area does not use press data (and therefore having a section within the press data section would not be appropriate), but also because much of my analysis in this thesis discusses directly and indirectly (e.g. through discussing prescribed linguistic forms) the role that language has in stigma creation and perpetuation.

In the next sections (2.3.1-2.3.2), I report on research on the theme of stigma and stigma reduction efforts. In Section 2.4-2.4.4 I report on research that uses press data. In Section 2.5 I provide an overview of the literature on stigma and the representation of mental illness in the news before I position my research in relation to the existing literature in Section 2.6. In Section 2.7, I conclude.

2.4.1 Stigma and self stigma

Perhaps the most widely researched area of the social construction of mental illness is the study of stigma. Stigma, Björkman et al. (2007) write, "has been identified as one of the most important obstacles for a successful integration of people with mental illness into the society" (2007: 332). The term was first discussed in detail by Goffman

(1963) in *Stigma: Notes on a Spoiled Identity*, and, as a result of Goffman's work into stigma, social scientists paid increased attention to the social and political consequences of it. As a result, it has been the focus of a significant body of research spanning psychiatry, sociology, psychology and linguistics. For example, Corrigan (1998: 202) writes that "the negative impact of severe mental illness is not entirely due to the ramifications of a biological disorder. Society's reaction to the disease seems to have an equally harmful impact on the person's abilities to successfully achieve life goals". The discrimination and prejudice that constitutes stigma surrounding mental illness has been termed a "psychosocial" (Wahl, 2012: 9) aspect of illness that is a significant obstacle in the process towards recovery. In a two-decade long review of mental illness stigma in the mass media, Klin & Lemish (2008) found that mass media representations of mental illness were "exaggerated, distorted, or inaccurate" (2008): 443) and that people with mental illness were presented as violent and dangerous. In addition to this, Corrigan & Watson (2002) found that stigma is not something that exists solely outside the individual with a mental illness, but also affects and plays a key role in the "personal power" (Corrigan, 2002) and self-esteem (Corrigan & Watson, 2002: 35) of individuals with a mental illness. As a result, "individuals with mental illness who then perceive negative actions by others to be legitimate will manifest lower self-esteem and diminished self-efficacy." (Corrigan & Watson, 2002: 47). Moreover, studies have discovered that even trained professionals have been found to "diminish the personhood of those labelled as mentally ill" (Lyons & Ziviani, 1995:1007). This means that due to the stigma surrounding mental illness it may be the case that people with a mental illness are not accessing services to help them due to a fear of being stigmatised, and professionals are not providing the services they ought to because they hold stigmatising views. More worrying still, research has shown that inaccurate media depictions of mental illness "could sometimes overwhelm direct experience in this area" (Philo, 1997: 171). Goodwin et al. (2016) echo this sentiment, and write that "Fictional depictions of mental health diagnoses may make for dramatic and entertaining viewing, but such inaccuracies present

audiences with misinformation and contribute to stigma." (2016: 385). However, stigma is not something that affects all mental illnesses in the same way. Research has shown that different mental illnesses are more or less stigmatised that others. For example, Mann & Himlein (2004) found that the "stigmatization of schizophrenia was significantly higher than stigmatization of depression". Moreover, in their participant sample of undergraduate students in America, they found there to be "significantly less stigmatization of mental illness […] among females than among males." (Mann & Himlein, 2004: 185). Dietrich et al, (2006) also found that there is a link between stigmatising media representations of people with a mental illness and negative attitudes towards people with mental illness (Dietrich et al., 2006).

In addition to stigma from others, there is a wealth of research into the negative effects of self-stigma on people with mental illnesses, specifically their willingness to perceive their symptoms as being to do with mental illness and also their willingness to engage with mental health professionals. Schomerus et al. (2012) found that people with a depressive syndrome were less likely to perceive their symptoms as being to do with mental health (what they term as 'lower problem appraisal') and as a result are less likely to seek the help of professionals if they held self-stigmatising attitudes. They concluded that personal stigmatising attitudes posed a barrier to seeking professional health care. Moreover, such lack of engagement with primary care and specialist mental health services has implications for the worsening of mental illness conditions.

Research into the stigma surrounding mental illness has been plentiful; however, all of the studies reported so far in this section presuppose the existence of stigma, i.e. they are all top-down analyses. This is not to say that stigma does not exist. I believe it does, but the effect of embarking on the analysis of stigma as an *a priori* construct means that the existing research into mental illness stigma focusses almost entirely on the effect of stigma, rather than how stigma is created and – specifically for the linguist interested in stigma – how stigma is created through language. For example, Corrigan & Watson (2002: 38) draw attention to the fact that stigma is the result of stereotypes,

prejudice and discrimination against people with mental illness but they do not unpack what these stereotypes are or how the vague concepts of prejudice and discrimination are actively enacted in different social settings. In another example of in the psychiatry literature concerned with stigma where the term is not fully explored, Thornicroft et al. (2007) state that "the term stigma refers to problems of knowledge (ignorance), attitudes (prejudice) and behaviour (discrimination)" (Thornicroft et al., 2007: 192). Again, this definition only provides vague descriptions of how stigma is manifest and provides no examples of how these concepts may be realised through language. Similarly, Kvaale et al. (2013) describe stigma as being related to the concepts of blame, dangerousness and social distance, but they too do not define what they mean by these terms or how the reification of these concepts may influence stigmatising attitudes. Furthermore, the methods employed to assess levels of stigma are limited in their scope. For example, Schomerus et al. (2012) based their findings on self-stigma entirely on the statistical analysis of Likert scale results where students with depression were asked to state the degree to which they agreed or disagreed with a set of statements. The statements the researchers used to elicit data were all negatively framed, e.g. "I believe people with mental illness are dangerous". As a result, their findings could be said to be biased towards negative results due to this negative framing. Moreover, findings that rely on methods such as Likert scales are limited in their scope because such methods are restricted with regard to how much information it is possible to amass from participants, given that the aim of Likert scales is to quantify opinions on something (and therefore the data collected is discrete). Moreover, research into the experimental design of Likert scales has shown that participants completing Likert scales can exhibit "end-aversion bias" (Hassan & Arnetz, 2005: 4), where participants avoid the extreme ends of the scale, which can result in skewed results. Additionally, researchers have argued that whilst measures such as Likert scales provide initial information about opinions on a topic, they are "not appropriate on a stand-alone basis" (Torrance et al., 2001: 329). Taken together, the existing research into stigma and self-stigma, then, presupposes the existence of

stigma whilst offering no or very little explanation of what constitutes stigma. Due to the assumption of stigma as an *a priori* construct, previous research also focuses on the effect of stigma rather than how it is created or reinforced. Furthermore, previous research into stigma is over-reliant on discrete models to elicit data such as Likert scales and questionnaires. As a result of elicited data using predetermined questions, the measures of assessing stigma are limited in scope.

In the next section, I review the existing literature into anti-stigma initiatives.

2.4.2 Stigma reduction efforts

In this section, I describe the existing research into stigma reduction, such as anti-stigma initiatives. Due to the fact that anti-stigma initiatives in the UK are few and still ongoing (and therefore the efficacy of them is still being assessed), I will also report research into the anti-stigma initiatives in other cultures.

As I discussed in the previous section, research into the stigma surrounding mental illness has reported that it has real-world consequences, such as making people with mental illness feel unable to access primary health care as well as stay engaged with that care. Furthermore, research has shown that there is a link between the stigma around mental illness and people's attitude towards those with a mental illness, including, in some cases, medical professionals and those with direct (and contrasting) experiences of people with mental illness. Furthermore, increased awareness of and education about mental illness results in decreased stigma (Simmons et al., 2017). Based on this body of research we can be fairly confident that stigma exists (even if the ways in which it is manifest are not discussed in the previous research) and that it has an effect on how people perceive mental illness and people with mental illnesses. In research since Goffman's work on the nature of stigma, researchers have explored ways in which stigma can be reduced by creating anti-stigma programmes and assessing the efficacy of national anti-stigma media campaigns. More recent research has discussed in detail the cultural differences

surrounding mental health stigma. For example, Fung et al. (2011) developed a self-stigma reduction program for 66 adults with schizophrenia in Hong Kong. They argue that stigmatisation of people with schizophrenia in China is greater than in Western societies due to the collectivist nature of Chinese culture, or the tendency to "place the need, interest, and objective of in-groups at a higher priority than that of the individual" (Wang & Chen, 2010). Following Lam, et al. (2010), Fung et al. (2011) argue that "under the collectivistic ideation, deviant behaviors of schizophrenia are regarded as character flaws or low moral standards, an interpretation that in turn results in higher levels of discrimination (2011: 208). Fung et al. (2001) found that whilst the program had short-term positive effects on the individuals, the effects were not long-lasting. In addition to Fung et al.'s (2011) findings from their stigma reduction campaign, a range of studies have demonstrated that TV and social media campaigns (hereafter SMCs) can change public attitudes about mental illness and the stigma associated with it (Søgaard et al., 1995). For example, Stuart (2003) explored the differences made by one Canadian newspaper before and after the pilot project of the global anti-stigma program '*Open the Doors*' (Stuart, 2003: 652). Stuart found that as a result of educating reporters about mental illness, the number of positive stories increased by a third and the overall length of the stories increased by a quarter (Stuart, 2003: 651). Moreover, in more recent and more broad research, Sampogna et al. (2017) assessed the effect of anti-stigma marketing campaigns in the UK between 2009-2014, during which time the anti-stigma SMC *Time to Change* had been in effect. Using questionnaires to assess attitudes towards mental illness in the community (n=10,526), they found that there was a positive correlation between awareness of the *Time to Change* SMC and higher scores on the questions concerned with tolerance and support (Sapogna et al., 2017: 116). As a result, they concluded that SMCs can have a positive effect on stigma reduction. This result built on previous research into the effectiveness of *Time to Change* conducted by Thornicroft et al. (2013), who found that there was an increase in positive articles after the *Time to Change* SMC with the number of stigmatising articles staying the same.

As was the case with the research I discussed in the previous section, the research into the efficacy of the various anti-stigma campaigns have been conducted by researchers working in the field of psychiatry. As a result of the fact that much of the previous research was conducted in one field, the methods used to research the efficacy of anti-stigma initiatives have many of the same methodological issues that I reported in the previous section. Namely, previous research into anti-stigma initiatives does not explore what the concept of stigma is in any great detail, but presumes its existence. Moreover in such studies, language (which presumably is a fundamental part of how stigma is manifest) is not an object of study. To give an example of how previous research neglects language analysis, I will revisit the research conducted by Sampogna et al. (2017), who explored the *Time to Change* SMC, because their methods are similar to many other studies. Sampogna et al. (2017) used three very commonly used questionnaires (sometimes referred to as scales) that are designed to assess participants attitudes, knowledge about and behaviour towards people with mental illness. Just like those reported in the previous section, the questionnaires are Likert scales. The questionnaires are titled the *Community Attitudes toward Mental Illness (CAMI)* questionnaire (Taylor & Dear, 1981), the *Mental Health Knowledge Schedule (MAKS)* questionnaire (Evans-Lacko et al., 2010) and the *Reported and Intended Behaviour Scale (RIBS)* questionnaire (Evans-Lacko et al., 2011) (see also Granello & Gibbs, 2016; Henderson et al, 2016; Wahl & Leftkowitz, 1989; Wolff et al., 1996; for other examples of research that uses some or all of these questionnaire to elicit data). Previous research like that of Sampogna et al. (2016) has used one or a combination of the three questionnaires to elicit data (data here constitutes the participants' responses to statements on the questionnaires). Combining multiple questionnaires in one study allows the researchers to cross-reference findings, e.g. explore correlations between certain views (e.g. the view that people with mental illness are a threat to society) and the level of knowledge about mental illness (e.g. a high awareness of issues concerning mental illness may correlate with increased tolerance of people with mental illness). In using these questionnaires to elicit data

about stigmatising views then, Sampogna et al. (2016) assume that certain attitudes and behaviours constitute stigma (this is in line with Thornicroft et al.'s (2007) definition of stigma discussed in the previous section) rather than assess whether (and if so, how) stigma is manifest in data. The first problem with research such as that conducted by Sampogna et al. (2016) is that, just like the previous research reported above, it assumes that stigma exists and that stigmatising attitudes and beliefs are accessible through questionnaires. This is not to say that the questionnaires are not useful or that stigma does not exist, however. The three questionnaires are useful in that they allow for the quantification of attitudes, behaviours and knowledge. This is turn means that correlations between attitudes and behaviours and knowledge can be explored. However, I argue that these questionnaires are limited when used as the sole method of data elicitation. This is because they can only reveal what a participant *thinks* they believe and *thinks* about how they would act rather than how they actually do. Moreover, due to the Likert scale design of the questionnaires, a participant's answers are not just based on their own potentially unreliable account of their thoughts and behaviours, but are also limited by the set number of responses they can give (e.g. dis/agreement words on a 5-point scale). In addition to these problems, the participants in many studies are students or professionals working in psychiatry or psychology, which may mean that they have a vested interest in answering the questionnaire in a certain way. To illustrate my point by way of an analogy from linguistics (where I use the pronunciation of a word as akin to an statement on a Likert scale), a sociolinguist would not ask a person how they pronounce a certain word because to do so would necessitate people being able to accurately reflect on their own pronunciation, which may be tied up in all sorts of ideology surrounding the prestige of certain accents (just like attitudes towards people with a mental illness are ideologically loaded), resulting in the person not pronouncing the word in the way they would unprompted. Furthermore, a sociolinguist would not ask a person how they pronounce a certain word because to do so could potentially bias the person's view of how they do in fact pronounce it, resulting in the participant accommodating

to the sociolinguist (just like a student answering a survey may exhibit participant bias). Moreover, how a participant in a sociolinguistic experiment pronounces a word may vary depending on the context, just as attitudes may change depending on the context. To explore variation in pronunciation, then, sociolinguists use various methods, such as asking participants to read passages aloud that feature a certain phoneme in different positions of a word. Likert scales cannot assess participants' attitudes outside of the hypothetical conditions described on the Likert scale questionnaire.

Notwithstanding these problems with Likert scale design (and the problems I discussed in the previous section related to end-aversion bias), the questionnaires are limited in what they reveal because, as I alluded to in my analogy, they are open to participant bias (e.g. where a participant answers in a certain way in order to fulfil what they perceive the purpose of the experiment to be – as in the sociolinguistics example). Also, there is arguably a greater risk of participant bias from the use of methods designed to test people's attitudes towards people with mental illness, simply because people may not wish to appear to hold stigmatising attitudes about such people. This is particularly the case in previous research that has used the *Community Attitudes to Mental Illness* (CAMI) scale on participant groups with a vested interest in people with mental illness such as counselling professionals (e.g. Granello & Gibbs, 2016).

Knowing this, I argue that perhaps a more revealing way to explore stigmatising attitudes and behaviours towards people with mental illness is to look at written data, e.g. newspaper articles. The reason for this is that language is one way that authors can encode ideology (i.e. stigmatising attitudes about mental illness) in a way that is unobvious, and in some cases, unintentional (and therefore revealing of unfelt ideology surrounding mental illness). Some research in psychiatry has explored the efficacy of anti-stigma initiatives using press data. For example, Rhydderch et al. (2016) explored the effect that the *Time to Change* anti-stigma initiative had on newspaper coverage from 2008-2014. They found that newspaper coverage of mental

illness has increased over time (Rhydderch et al., 2016: 45). In their analysis, the research team identified the central theme of each newspaper article for any 'element' (roughly akin to topic, e.g. 'recovery and treatment', 'danger to others') which was "stigmatising, antistigmatising or neutral." (Rhydderch et al., 2016: 47). Rhydderch et al. (2016) do not state what constituted a central theme, or what criteria they used to assess whether a given 'element' of the text was stigmatising, antistigmatising or neutral. They do however, list "pejorative language" as an element of interest (Rhydderch et al., 2016: 47). They do, then, view language as contributing to stigma, but they do not offer any explanation of what they see pejorative language to be. For this reason, whilst Rhydderch et al.'s (2016) research does recognise the role language plays in stigma creation, it is not replicable because the parameters of what stigmatising elements are, or what constitutes pejorative language are not specified.

With these problems in mind, a systematic analysis of large quantities of language data on the topic of mental illness that uses well-documented and replicable analytical methods, such as those of corpus linguistics, provides a better means of analysing stigma in newspaper discourse. Moreover, computational analysis of such data (like that conducted in corpus linguistic analyses) means that common linguistic patterns, which may relate to stigma, are more easily identified than in predesigned questionnaires on stigmatising attitudes. Furthermore, due to the fact that linguistic analysis of newspaper discourse is conducted after the texts are written, the participant bias I discussed previously is not an issue.

In the next section, I review the existing research into the representation of mental illness that uses press data.

2.5 Press data

A wealth of research has been conducted on representations of mental illness in the press. The reason for press data being so widely used in studies is that it is still a major source of information for the public about mental illness (Philo et al., 1994). For

example, Nawková et al. (2012: 22) write that "even in the era of the internet, printed media are still among the most frequently identified sources of mental health information". The studies conducted of mental illness using press data have many different focuses; for example, diachronic studies to assess attitudes to mental illness over time, gender differences in the reportage of mental illness, depictions of mental illness and criminality and representations of specific illnesses. In section 3.3.1 I will provide an overview of the key themes in existing research. In the sections following 3.3.1, salient themes in the research will be discussed in more detail.

2.5.1 Mental illness in the news

As was demonstrated in Section 2.3, a vast proportion of research into mental illness has focussed on the stigma of mental illness and negative portrayals of mental ill health in the press. Stuart (2003) writes that "negative media stereotypes are among the most hurtful and socially limiting stigma experiences reported by mental health service consumers and family members." (2003: 651). Despite research reporting that press coverage of mental illness is getting better (Francis et al, 2004; Whitley & Wang, 2016), other research shows that there is still a focus on mental illness over the positive aspects of mental health such as wellbeing (Kenez et al. 2015). Consistent findings in much of the existing research is that mental illness is reported often, is highly topical (Ohlsson, 2017), and is reported as an epidemic (Ohlsson, 2017: 309; Bilić & Georgaca, 2007). Furthermore, research has consistently shown that news articles on mental illness rarely feature any report or quote from a person with a mental illness. For example, Nairn & Coverdale (2005) found that only 0.8% of 600 newspaper articles taken from New Zealand newspapers featured any self-depiction or report from a person with a mental illness. This supports earlier research which noted an absence of voices from people who have first-hand experience with mental illness in media reporting (Wahl et al., 2002). More recent research conducted in Bermuda found that over a 20-year period the number of mental health professionals quoted had increased

while articles featuring servicer users stayed the same. Moreover, where a service user was quoted, the quote was usually taken from court reports pertaining to violent crime (Roberts et al., 2013: 388).

Reporting mental illness in articles that focus on violence, serious crime and dangerousness is also a major cross-linguistic theme in media reports on mental illness. (This is despite evidence that suggests that age, gender and ethnicity are more accurate predictors of violent crime than severe mental illness (Fazel & Grann, 2006)). For example, in a survey of three European countries (Czech Republic, Slovakia and Croatia), Nawková et al. (2012) found that news stories relating to mental illness varied across the countries but were consistently negative across the three countries. They also found that positive articles were more likely to be around 50% longer than negative ones, leading the researchers to say that "longer articles are more positive because the journalist has more room to give accurate details" (Nawková et al., 2012: 7). Moreover, they found that despite statistics showing that people with mental illnesses are more likely to be victims of aggression, very little attention was given to this topic (2012: 8). They also found that articles did not cover issues to do with recovery and therapy that showed the positive aspects of living with mental illness (2012: 8). The tendency to report mental illness sensationally and in reference to violence and aggression in the media has led to concerns that such coverage may create a "moral panic" (where a "group of persons emerges to become defined as a threat to societal values and interests" (Cohen, 1973: 2)) around mental illness (I discuss the notion of moral panic in Chapter 7). Such research into moral panics includes that by Paterson (2006) who explored the association between violence and mental illness. Furthermore Hallam (2002) explored the link between violence and mental illness in reference to community care policy (Hallam, 2002).

Rose (1998) also explored the link between the perceived failures of community care (as compared with institutional care) through an exploration of coverage of violent crimes in the media, notably the case of Christopher Clunis who stabbed and killed Jonathan Zito in London in 1992. Clunis had schizophrenia and had been

discharged from hospital prior to the attack. Reports after Zito's murder focused on how community care had failed in treating and Clunis and preventing such an attack. Rose's study echoed the findings of Thornton & Wahl (1996) who found that after reading a news article reporting on a murder committed by a person with mental illness while on day release from a psychiatric hospital, participants were more likely to view people with mental illness as in need of "monitoring and restriction". (Wahl, 2003: 1596). This, Wahl writes, "fuels resistance to community care" (2003: 1596). Thornton & Wahl's findings are based on an experiment in which students in an introductory psychology class were randomly given one of three experiment packs each containing a newspaper article that Thornton & Wahl deemed stigmatising and one other article which was either a "fact-orientated" article, an article that "addressed misconceptions about mental illness and gave correct information on mental illness", and an article that "discussed media distortion of mental illness" (Thornton & Wahl, 1996: 18-19). The experiment also included the use of a fourth pack that did not include any article deemed to be stigmatising. This was the control condition. The students were then asked to fill in series of questionnaires, one of which was the CAMI scale I described in the previous section. Whilst the findings that Thornton & Wahl (1998) report are compelling (i.e. that there is a link between media depictions and the view that people with mental illness should be monitored and restricted), the findings are predicated on methods that are limited for the reasons I have previously outlined. These reasons are that (i) the authors do not specify the characteristic features of the text that they say is stigmatising (above the fact that the text features a violent event), (ii) they use questionnaires to elicit the participants' opinions only, and (iii) because their participants are psychology students, they are not representative of the general public. Furthermore, the experiment does not accurately reflect the means by which people consume newspaper articles; e.g. it is highly unlikely that a member of the public would read two articles on mental illness in such close proximity that one would affect the interpretation of the other in such obvious ways. The creation and

reproduction of stigma in the news is much more pernicious than Thornton & Wahl's experiment suggests.

Thornton & Wahl's experiment is limited too in that the number of articles presented to participants was very small (just two). In a larger scale study, Wahl et al. (2002) analysed the news coverage of mental illness in 300 newspapers. However, this research was still based on the subjective analysis of themes in newspaper reports. For example, Wahl et al. (2002) describe their method as including the identification of "the main themes of each article" and "the overall tone of the article" (Wahl et al., 2002: 9). As a result, the study is limited in its replicability because thematic analysis of this kind relies on the interpretation of the individual analyst. Wahl et al. (2002) do, however, make reference to language in their research. They analyse the frequency of person-first language (e.g. 'people with mental illness' rather than 'the mentally ill'). This recognition of the role language plays in creating stigma is positive; however, the research team refer to person-first language without fully exploring whether there is a linguistic basis for person-first forms.

In their research, Wahl et al. (2002) found that very little attention was paid to community intervention in the newspaper coverage. Reporting on this earlier work, Wahl (2003) writes:

> Seldom were psychosocial or community interventions described, despite their increasing importance in the recovery movement. Thus, current newspaper coverage contributes to the medicalization of mental illness and the public is led to accept—and provide financial support for—medication and hospitalization as the primary solutions for mental health problems. Similar acceptance and support for psychosocial intervention, rehabilitation, and community treatment will likely be harder to obtain given their absence from journalistic considerations of mental health treatment options.
>
> (Wahl, 2003: 1598)

Media reportage, then, can affect attitudes towards people with mental illness in a local sense (i.e. stigmatising an individual), but also in a national sense as moral panics caused by media reports may influence social policy, such as a movement away from community care and coercive mental health policy.

There are other trends in the research into mental illness in the news. For example, the coverage of causes and responsibility for mental illness. Zhang et al. (2016) found that historically the responsibility for and the causes of depression were more likely to be attributed to the individual than to society in US newspapers, writing that "the number-one causal factor presented by the media was genetics, personality, and individual health outcomes" (2016: 128). Moreover, Corrigan et al. (2005) found that attributing the cause of mental illness to environmental factors was more common than attributing the cause to individuals, and Ohlsson (2017) found that an increase in mental health problems in the Swedish print media was being attributed to life in modern society (2017: 302). As a result of this, Ohlsson states that often the concept of mental health is taken for granted in the news reports and, because of this, "there is a conceptual confusion when it comes to what should be regarded as medical problems and what should rather be seen as other kinds of painful experiences." (2017: 304). Ohlsson (2017) argues that mental illness is used in a vague way to describe specific mental illnesses (e.g. schizophrenia) to emotional (i.e. non-pathologised) pain). Confusion over what constitutes mental illness was also a theme found by Pescosolido (1999) who found that members of the American public were less likely to view depression as severe or, in some cases, as a mental illness at all. (1999: 1343). More recently Rowe et al. (2003) found that the reporting of depression in Australian newspapers often neglected to provide any definitions or explanations of depression and that depression was a taken-for-granted concept. They also found that depression was often compared with "other, genuinely biological, illnesses […] where medico-scientific language is used to account for lived experience" (Rowe et al., 2003: 692); i.e. experts were promoted rather than members of the public.

There also exist trends in the coverage of particular participant groups in newspaper reports on mental illness. For example, Slopen et al. (2007) analysed news articles to assess mental illness reportage and responsible journalism. The research team use the term 'responsible journalism' to refer to those newspaper articles that adhered official media guidelines set out by mental health professionals in Australia (e.g. the guidelines included avoiding "slang terminology"). They found that articles related to children were more likely to refer to behavioural issues and alcohol or drug abuse while stories on adults were more likely to refer to crime and dangerousness. They also found a greater degree of responsible journalism in the child stories, whereas stigmatising language was more likely to be found in adult stories (Slopen et al., 2007: 3). Slopen et al.'s (2007) findings are interesting because they are revealing of the different journalistic practices in reports on mental illness; however, the methodology used does not query the role that language plays in any detail. They never unpack what the term 'slang' refers to; they simply list some lexical items that they coded as slang. This list includes words such as "psycho" and "nuts", which may be conventionally associated with slang use, but they also list the words "lunatic" and "madness", despite the fact that there is no basis for defining these as slang terms (especially in the UK where these terms would have been used in official mental illness legislation historically, e.g. *The Lunacy Act* of 1845). Moreover, Slopen et al. (2007) study slang usages with the assumption that such usages are inherently stigmatising. Doing this is clearly problematic from a linguistic perspective where the context of any utterance (written or otherwise) has to be taken into account in order for something to be deemed as offensive, and so no work is inherently stigmatising. Further to Slopen at al.'s (2007) finding that there are different journalistic practices depending on who the person with mental illness is (i.e. adult or child), Coverdale et al. (2002) found that in newspaper articles in New Zealand, more articles concerned male mental illness than female mental illness with no articles on child or adolescent mental illness.

A further theme that is covered in much of the existing literature is whether or not articles refer to recovery or treatment. Corrigan et al. (2005) found that there was equal representation of biological and psychosocial treatments in a sample of US newspaper reports on mental illness (n=3353) with just 4% of those addressing recovery (Corrigan et al., 2005: 551). More recent articles have shown that recovery is a theme in news articles reporting on mental illness. In one of very few studies into press representations of mental illness that is informed by linguistics, Atanasova et al. (2019) found that recovery was the most prominent theme in a corpus assisted analysis of 1,412 British newspaper articles. Moreover, in contrast to the prevailing trend in psychiatry research that assumes that reports on mental illness are stigmatising, Atanasova et al. (2019) found that the discussion of stigma (e.g. raising awareness about it) was a key theme in newspaper reports.

In the following sections, specific themes in media depictions of mental illness research will be covered in more detail.

2.5.2 Depictions of criminality and violence

So far in Section 2.4, I have reported on the existing literature that has found an association between mental illness and violence and criminality. In this section, I explore research in more detail. The reason for doing this is that research into depictions of criminality and violence in reference to mental illness have been numerous. Consequently, this topic warrants its own section.

As I have shown in my report on the existing literature so far, mental illness is consistently presented negatively in the media, which results in stigma around mental illness. One of the ways in which this stigma is created is through the reporting of mental illness alongside reports of criminality or violence. For example, Bowen (2016) found that 42% of UK newspapers that reported on personality disorder between 2001-2012 linked personality disorder and homicide (2016: 601) with articles that linked personality disorder and homicide decreasing in the later time period (2007-

2012). Bowen's findings echo those of Whitley & Berry (2013) who found that 40% of Canadian newspaper articles on mental illness between 2005-2010 (n=11,263) were related to violence & criminality. Additionally, what is and is not included in reports on mental illness can result in incorrect assumptions about the symptoms of illnesses. For example, Vilhauer (2015) explored depictions of auditory verbal hallucinations (hereafter AVH) in US newspaper data (n=181) and found that the media reported AVH as a pathology rather than something that can occur in psychologically healthy people. Vilhauer (2015) argues that the representation of AVH as a pathology could increase stigma around AVH. Moreover, Vilhauer (2015) found that people who experience AVH were portrayed negatively, with AVH being associated with violence and criminal behaviour (2015: 61) (see Deamer & Hayward, 2018 and Demjen et al., 2019 for linguistic accounts of voice hearing using non-press data).

Looking specifically at the link between violent crime and mental illness, Flynn et al. (2015) explored the reporting of 60 homicide-suicide cases in newspapers in England and Wales over a three-year period. They found that pejorative and derogatory language (terms that they do not unpack but rather assume are obvious) was used in reference to mental illness with the depiction of mental illness being inaccurate. They found several themes (guided by an existing and generic qualitative analysis framework devised by Hodder 2010 [2003]) in the articles analysed (n=1163) including blaming, violence and personal tragedy, mental illness and speculation that the incident was due to mental illness (Flynn et al., 2015: 270). In line with these findings, McGinty et al. (2016) found that in a sample of US newspapers published between 1995-2014 (n=400), there was an increase over the time period in articles that mentioned mass shootings committed by people with mental illnesses (2016: 1121). Furthermore, Whitley et al. (2017) explored Canadian newspaper articles that reported on cases where a person has been found not criminally responsible on account of mental disorder (NCRMD) compared with general articles on mental illness. They found that articles containing references to NCRMD were more negative

overall and "almost never focused on recovery or rehabilitation, in stark comparison to generic articles about mental illness" (Whitley et al. 2017: 697).

Previous research has also shown that reports on certain illnesses are more likely to feature depictions of people with mental illness as criminal or aggressive. For example, Nawka et al. (2012) analysed 375 Czech Republic and Slovakian newspaper articles from 2007. They found that 31.2% of the data reported aggressive behaviour, with homicide being most frequently mentioned in reference to psychotic disorders and schizophrenia whereas suicides and homicides were more frequently reported in reference to affective disorders. Additionally, they found that eating disorders and anxiety disorders were not linked to any aggressive behaviour (Nawka et al., 2012: 1). Additionally, Coverdale et al. (2002) found that in a sample of newspaper articles in New Zealand, a key theme was criminality, but also vulnerability (the notion that people with mental illness are "incompetent and unable to control their own life" (2002: 699). They also found that in positive articles, common themes were human rights, leadership, sporting prowess or educational accomplishments (2002: 699). More recent diachronic studies have demonstrated a change in media representations of mental illness towards more positive reporting. For example, Goulden et al. (2011) used content analysis to analyse UK newspaper articles in three different years (1992, 2000, 2008) (n=1361). They found there were fewer negative articles over the time period with an increase in articles on psychiatric disorders featuring explanations of this category of mental illness. However, the coverage of illness types was variable. They write:

> The overall positive trend masks considerable variation by diagnosis. The reporting of depression, anxiety, bipolar disorder, and eating disorders, either improved over time or was always largely favourable. In contrast, schizophrenia, personality disorders, and general references to mental illness, appeared mainly in the context of 'bad news', and saw little or no change in their coverage over time.
>
> (Goulden et al. 2011: 5)

These findings are echoed by Whitley & Wang (2016) who analysed articles that discussed mental illness from Canadian newspapers between 2005-2015 (n=24, 570). They found that articles with a positive tone had doubled over the time period and stigmatising content had reduced by a third (2016: 278). They argue that these figures suggest that national anti-stigma campaigns have been successful.

The research I have reported in this section exhibits a fairly standard methodological approach to the analysis of press data in the existing literature. This consists of collecting newspaper articles based on arbitrary search terms, conducting some form of thematic or content analysis (e.g. Coverdale et al., 2002; Goulden et al., 2011) with the aim of identifying whether the articles are positive or negative (e.g. Goulden et al., 2011; Whitley & Wang, 2016) or identifying key themes (Coverdale et al., 2002). The problem this methodological approach is that whether an article is viewed as positive or negative, or what constitutes a key theme, relies heavily on the analyst's own interpretation of the data, which makes the research hard to replicate (recall that I argued this same point about Rhydderch et al.'s (2016) research into anti-stigma initiatives).

2.5.3 Gender and mental illness

There are a number of studies exploring gender differences in reports on mental illness. The reason that gender is an interesting variable to study in mental illness reportage is that some mental illnesses are more prevalent in certain genders. For

example, unipolar depression is twice as common in women, and men are "more than three times more likely to be diagnosed with antisocial personality disorder than women" (WHO, 2019). Moreover, how a society socially constructs an illness can create stereotypes about who can be affected by specific mental illnesses. For example, there is a stereotype that only women get eating disorders and that depression is at odds with a masculine identity (Galasiński, 2017). Previous research has also shown that the mass media present gender biases in relation to mental illness; for example, Klin & Lemish discuss how women's magazines frame stress and agoraphobia as "female mental disabilities" (2008: 438) whereas magazines over-represented men in discussions about "psychoses, personality disorders and childhood problems" (Klin & Lemish, 2008: 438).

Research has shown that men do not access help for mental illness (Johnson et al., 2012) because seeking help for depression is seen as "culturally feminized" (Kilmartin, 2005; Scholz et al., 2014). Machlin et al. (2014) explored the link between positive portrayals of men with depression in the news and men seeking help for depression and anxiety via helpline services in Australia. They found that there was an increase in the uptake of helpline support in the weeks after the positive news stories. Furthermore, Whitley et al. (2015) explored gender in mental illness reports. Using content analysis to analyse 1168 newspaper articles collected over a six-month period, Whitley et al. (2015) found that newspaper articles about men were more negative than those about women. Whitley et al.'s (2015) study is another example where the method used is heavily reliant on interpretation, e.g. content analysis. In a description of their method, Whitley et al. (2015) write that they "read and coded articles for the presence or absence of themes and content" (Whitley et al., 2015: 325) but offer no description of what constitutes a theme. As a result of the overreliance on interpretation and unclear parameters of what constitutes a theme, Whitley et al.'s (2015) research is not replicable.

Finally, in research into suicide, Marzano et al. (2018) found that in a collection of 8,101 articles published in the UK and the Republic of Ireland over a 12-month

period; reports featuring young females (particularly those whose suicide was considered unusual) were more common.

Existing research, then, has shown clear gender differences in the reportage of mental illness.

2.5.4 Depictions of schizophrenia

The reason for having a separate section on studies into newspaper representations of schizophrenia specifically is that research into schizophrenia in the press is hugely overrepresented in the literature to date. This is intriguing when one considers that newspaper articles on schizophrenia are less common than those discussing other illnesses (Wahl, 1996). A considerable amount of research into the representation of schizophrenia in print media has been concerned with the label itself, which is medically contested, even leading some mental health professionals to state that schizophrenia does not exist (van Os, 2016).

An example of research in this area is Lasalvia et al. (2015), who conducted a review of research into the label 'schizophrenia' in light of mental health professionals and researchers calling for the term to be abandoned, arguing that it was stigmatising (Howe et al., 2014; Read et al., 2006). Lasalvia et al. (2015) analysed 47 articles published worldwide on the topic of renaming schizophrenia and new candidate labels. The team found that, overall, the research suggested that relabelling the condition would be a positive step, as non-medical usages of *schizophrenia/schizophrenic* "make it difficult to convey its proper meaning and contribute to maintaining the misconceptions and misinformation about the condition" (Lasalvia et al., 2015: 282). The use of *schizophrenia* or *schizophrenic* to refer to things outside of a medical context has been a source of much research. Studies have also shown that in contrast to other mental illnesses, schizophrenia reportage is

more negative overall[9] (Aoki et al., 2016, Thornicroft et al., 2013; *inter alia*). One area of research into media depictions of schizophrenia that has attracted much attention is the use of schizophrenia as a metaphor (Duckworth et al., 2003; Guarniero et al., 2017; Lampropoulos et al., 2017, *inter alia*)[10], although these studies have tended to be carried out by researchers working outside of linguistics and therefore the categorisation and systematic analysis of source/target pairings are neglected, with statistical information about metaphorical usage being reported instead. Moreover, whether or not these studies refer to metaphor as it is known in linguistics and the cognitive sciences, or whether metaphor in the studies refers to any non-medical usage is unclear, as the reported method for analysis is often focused on the quantitative coding process rather than the linguistic instantiations of metaphorical language. Nevertheless, these studies do offer interesting insight into how often schizophrenia is used to refer to things outside of a medical context. Some researchers have categorised metaphorical usages. Frequent usages of *schizophrenia* as a metaphor in these studies include *schizophrenia* to refer to "split personality" or unpredictability. For example Magliano et al. (2011) found that metaphorical usages of *schizophrenia* were more common than references to it as a medical illness, with the metaphorical usages being used to refer to incoherence (related to the split personality metaphor), dangerousness/aggressiveness and eccentricity/oddness. (Magliano et al., 2011). These categories were also found in Greek newspapers (Athanasopoulou & Valimaki, 2014). Further to this, Duckworth et al. (2003) found that 28% of a sample of US newspapers (n=876) featured *schizophrenia* as a metaphor, which they state creates imagery that "encourages further stigmatization and a popular orientation that discourages individuals from seeking treatment for their illness" (Duckworth et al. 2003: 1402). In addition to this, Chopra & Doody (2007: 423) found that *schizophrenia*

[9] Aoki et al. (2016) focus on Japanese newspaper coverage and use the representation of bipolar disorder as a control condition.

[10] These studies often report Sontag's (1996) work on illness as metaphor and seem to be based on this view of metaphor.

was more likely to be used in a metaphorical sense than *cancer* was, but the reverse was true in US newspapers. Moreover, they report that 11% of their sample newspapers used *schizophrenia* as a metaphor.

Wahl (1996) looked at the representation of schizophrenia in the news over a 5-year period (1989-1994) in three American daily newspapers (n=101). Wahl (1996) found that the majority of reports pertained to the treatment of schizophrenia with drugs, or to the incidence of schizophrenia. Wahl (1996) found that the reportage was largely accurate. However, around 10% of the articles reported on schizophrenia in relation to criminal acts of a violent nature. Of the articles, 14 pertained to individual people's experiences with schizophrenia, which Wahl (1996) notes were typically sympathetic and presented the individuals positively. Wahl concluded that, overall, schizophrenia was underreported (however it is unclear whether Wahl is referring to articles that use *schizophrenia* in a medical sense or in a metaphorical sense).

In a later study, Angermeyer et al. (2005) explored whether there was a link between media portrayals of schizophrenia and stigmatising attitudes by assessing whether media consumption was related to the desire for social distance from people with schizophrenia. They found that participants who read tabloid and regional newspapers "express a higher preference for social distance towards people with schizophrenia than respondents who regularly read broadsheets (alone or in combination with either tabloids or regional newspapers or in combination with both) and those who do not read any newspaper at all." (Angermeyer et al. 2005: 248). Angermeyer et al.'s research suggests that there is a link between the media and stigma (and therefore stigmatising attitudes can be changed as a result of changing information in the media). Their findings also indicate sites to target as part of anti-stigma campaigns (2015: 249). They write that in order to affect positive change in relation to media depictions of mental illness "inaccurate and unfavorable messages should be replaced' by accurate and positive messages. Stories of people living with mental disorders should become commonplace in media reports." (Angermeyer et al. 2005: 249).

Pingani et al. (2018) conducted a diachronic study of the use of *schizo* in Italian press between 2001-2015. Their research aims were to survey articles pertaining to schizophrenia to "identify possible predictors reinforcing negative stereotypes about people with schizophrenia" (Pingani et al., 2018: 792). The researchers grouped the articles into usages of *schizo* in a medical sense and usages of *schizo* in a non-medical sense. They found that of the 946 articles analysed, 356 "mainly reinforce negative stereotypes regarding mental illness" (2018: 792). They found several 'predictors' that an article would contain negative stereotypes: "unnecessarily dramatic or sensational headline or content; inaccurate or not in the correct context use of medical terminology; emphasis to the illness rather than to the person; mental disorders are the same; disclosure of particular individual has a mental illness." (2018: 792). The researchers found that there was no reduction of negative stereotypes over the time period in articles that used *schizo* in a non-medical sense. Conversely they found that in articles that used *schizo* in a medical sense there was a significant reduction in stigmatising features. As the researchers point out, this finding is in contrast to studies conducted by Clement & Foster (2008) who found no significant change in the number of stigmatising articles on schizophrenia published in 1996 or 2005 in the UK press. They did find, however, that articles published more recently were less likely to use 'schizophrenia' in metaphorical contexts (2008: 178). The differences in the findings of Clement & Foster (2008) and (Pingani et al., 2018) arguably provide evidence for the cultural basis of illness that I discussed in Section 2.2 on social constructionism.

Gwarjanksi & Parrott (2018) conducted a content analysis of stigma frames to explore the representation of schizophrenia in articles taken from eight US news websites in 2015 (n=558). Unlike some previous research, Gwarjanksi & Parrott (2018) do describe what features were coded as positive and negative, although the descriptions are still vague overall, e.g. whether an article featured negative statements about people with mental illness, etc. Gwarjanksi & Parrott (2018) found that in keeping with research on print media representations of schizophrenia, the portrayals were negative (2017: 959). The researchers also found that reader comments

on the articles in which stigmatising frames were used were more likely to be stigmatising. In contrast, articles that contained stigma-challenging frames were more likely to feature reader comments that challenged stigma. This finding offers further evidence that the media can influence perceptions of mental illness.

Focusing on the difference in labels for the concept of schizophrenia, Aoki et al. (2016) explored whether there was a difference in the way schizophrenia was represented before and after the illness was renamed in Japan using data from three national broadsheets. In 2002, the name for schizophrenia was changed from "'seishin-bunretsu-byo', which literally means "mindsplit-disease" […] to 'togo-sitcho-syo', which literally means 'integration disorder'." (Aoki et al., 2016: 193). The researchers found a decline in articles that linked schizophrenia and danger after renaming when compared with a control condition. However, Koike et al. (2016) analysed newspaper headlines and discourse taken from a TV national TV programme before and after the name change and found that, contrary to Aoki et al.'s findings, schizophrenia was still reported in reference to criminality and violence (2016: 558). The difference in the findings made by Koike et al. (2016) and Aoki et al. (2016) may be down to the genre differences associated with the headlines of newspaper articles and the main body of the text, e.g. headlines are more sensationalised because the purpose of them is to be attention-grabbing (I discuss the features of headlines in more detail in Chapter 7) in addition to the fact that Aoki et al. (2016) use data from broadsheet newspapers only. Diachronic studies into the changing labels for schizophrenia and the effect of these changes such as Koike et al. (2016) have also been conducted using newspaper data from South Korean (Park et al., 2012) and in Chinese newspapers on psychosis (Chan et al., 2016), with few changes observed.

Up to this point, I have described the content of previous research into mental illness that uses language as data. Previous research has shown that that mental illness reportage is overwhelmingly and cross-culturally negative, with the levels of stigma associated with illnesses varying across illness types. Moreover, previous research has

showed that the language used to discuss and report on mental illness has real world consequences; for example, it has the potential to influence social policy and override the personal experiences of people with mental illnesses. Language, then, is central to our understanding of mental illness and the experience of mental illness.

My report to this point has described the findings of existing research, and has discussed some of the problems associated with the methods used in existing research. Table 2.1 details the methodological information for each study outlined in the press data section of this chapter. Presenting the research in this way allows for a clearer view of trends in existing research. Column 1 (C1) details the author and publication date, C2 details the main method or approach described in the method section of the article, C3 lists the number of newspapers that articles were collected from, C4 lists any particular focus of the article ('-' indicates no particular focus other than mental illness), C5 details the dates the data was collected from, C6 details the number of articles analysed, C7 details the country the data was collected from and C8 details the discipline of the journal the article was submitted to, and the subject area of the first named researcher. The reason for collecting the information listed in C8 is to provide insight into the disciplinary tradition the article belongs to (albeit by a rather blunt instrument). For example, in cases where discourse analysis is listed as a method, whether the author is from a linguistics background or a psychiatry background may mean the analysis is quite different in each case.[11]

[11] Table 1 is not exhaustive and is limited by what I have access to. I have endeavored to cover as wide a range of articles as possible using the Web of Science tool to assist my search.

OVERVIEW OF EXISTING LITERATURE METHODOLOGICAL INFORMATION							
Authors/Date	**Method/Approach**	**No. of sources**	**Specific focus**	**Dates covered**	**Sample size (n=)**	**Country**	**Journal discipline/lead author subject**
Aoki et al. (2016)	Content analysis/Pearson's correlation analyses, fishers r-to-z transformation	3	Schizophrenia & bipolar, stigma	1992-2012	4677	Japan	Schizophrenia/medicine
Atanasova et al. (2019)	Corpus-assisted frame analysis, thematic analysis	N/S	Arts initiatives	2007-2015	1,412	UK	Health/linguistics
Bilić & Georgaca (2007	CDA	~10	-	2004-2004	165	Serbia	Psychology /psychology
Bowen (2016)	Content analysis	6	Personality disorder	2001-2012	552	UK	Mental health nursing/health & social care
Chopra & Doody (2007)	Metaphor identification (using Sontag, 1996)	6	Schizophrenia & cancer	Aug 2004 – Nov 2005	600	UK	Medicine/practitioner
Clement & Foster (2008)	Content analysis	5	Schizophrenia	1996, 2005	1196	UK	Schizophrenia/primary care/public health
Corrigan et al. (2005)	Bespoke coding scheme (developed using Wahl, 2002)	70	Stigma	6 week periods every 2 months in 2002	3353	US	Psychiatry/psychiatric rehabilitation
Coverdale et al. (2002)	Thematic analysis	N/S	-	4 week period in 1997	600	New Zealand	Psychiatry/medical & health sciences
Duckworth et al. (2003)	Content coding	5	Schizophrenia & cancer	1996-1997	1,740	US	Psychiatry/N/S
Flynn et al. (2015)	Thematic analysis	N/S	Homicide-suicide	2006-2012	16,323	England & Wales	Psychiatry, medical & human sciences

Francis et al. (2004)	Bespoke instrument measuring responsible reporting	N/S	-	March 2000-Feb 2001	4351	Australia	Psychiatry/ Population Health,
Goulden et al. (2011)	Content analysis	4	Stigma	1992, 2000, 2008	1361	UK	Public health/community mental health
Guarniero et al. (2017)	Content analysis	1	Schizophrenia	2008	184	Brazil	Psychiatry/psychiatry
Gwarjanksi & Parrott (2018)	Content analysis	1	Schizophrenia	2001-2015	946	Italy	Psychiatry/psychiatry
Hallam (2002)	Non-specified	N/S	-	1992-2000	675	UK	Psychiatry/ economics of mental health
Kenez et al. (2015)	Content and thematic analysis	3	-	12 weeks in 2012	255	Australia	Public health/ psychology & public health
Koike et al. (2016)	Test data mining analysis	4	Schizophrenia	1985-2013	23169092	Japan	Schizophrenia/mental health
Lampropoulos et al. 2017	Intuitive coding scheme	8	Schizophrenia	2015	-	France	Social psychiatry/
Machlin et al. (2014)	N/S	N/S	Depression (male)	July 2012 – June 2013	10	Australia	Official report
Magliano et al. (2011)	Metaphor identification (using Sontag, 1996), v2	22	Schizophrenia	2008	1087	Italy	Social psychiatry/psychology
Marzano et al. (2018)	Bespoke coding scheme (for responsible journalism)	N/S	Suicide	12 months	8,101	UK and ROI	Crisis/psychology
McGinty et al. (2016)	Content analysis	7	-	1995-2014	400	US	Health/mental health & addiction
Nairn & Coverdale (2005)	Close reading Propositional analysis	N/S	-	4 week period in 1997	600	New Zealand	Psychiatry/population health
Nawka et al. (2012)	Content analysis	6	Aggression	2007	375	Czech republic & Slovakia	Psychiatry/psychiatry

Nawková et al. (2012)	Content analysis	6	-	5 week –long periods in 2007	450	Slovakia, Czech Republic, Croatia	Health communication/ psychiatry
Ohlsson (2017)	Thematic analysis	2	-	2009	691	Sweden	Health/ education
Ottewell (2017)	Content analysis	4	-	1987-2014	448	Japan	Public mental health/social sciences
Park et al. (2012)	Content analysis	3	Schizophrenia	2001-2010	490	South Korea	Social psychiatry/psychiatry
Paterson (2006)	Frame analysis	6	-	1992-2000	N/S	UK	Psychiatry/ nursing/midwifery
Philo et al. (1994)	Content analysis	N/S	-	April 1993	562	Scotland	Health education/ media
Pingani et al. (2018)							
Roberts et al. (2013)	Content analysis	4	-	1991, 2001, 2011	277	Bermuda	Psychiatry/psychiatry
Rowe et al. (2003)	Discourse analysis	N/S	Depression	2000	49	Australia	Sociology/psychology
Rhydderch et al. (2016)	Content analysis Coding system	27	Anti-stigma	2 days of every month between 2008-2014 (minus 2012)	N/S	England	
Scholz et al. (2014)	Discourse analysis	N/S	Depression (male)	Sept 2002-Aug 2011	849	Australia	Qualitative health research/applied psychology
Slopen et al. (2007)	Coded using bespoke system	-	Age analysis	1-week periods in 2002 every 2 months	1253	US	Health communication/ society & health
Stuart (2003)	Content analysis	1	Schizophrenia	24 months (year N/S)	~362	Canada	Psychiatry/ community health & epidemiology
Thornicroft et al. (2013)	Content analysis	27	Anti-stigma	2 days each month	~3000	England	Psychiatry/ health service &population research

				throughout 2011			
Thornton & Wahl (1996)	Questionnaire Experimental	-	-	-	3	US	Psychology / N?A
Wahl (1996)	N/S	3	Schizophrenia	1989-1994	101	US	Psychiatry/psychology
Wahl et al. (2002)	Thematic analysis, coded using bespoke system	N/S	-	1989-1999	600	US	Book/N/S
Whitley & Berry (2013)	Trend analysis	N/S	-	2005-2010	11,263	Canada	Psychiatry/psychiatry
Whitley & Wang (2017)	Trend analysis, chi-square test	~20	Anti-stigma	2005-2015	24,570	Canada	Psychiatry/ Psychiatry
Whitley & Wang (2017)	Thematic analysis	20	Criminality	2015	940	Canada	Psychiatry/psychiatry
Whitley et al. (2015)	Content analysis	N/S	Chivalry hypothesis	Nov 2010-apr 2011	1168	Canada	Psychiatry/mental health
Yang & Parrott (2018)	Quantitative content analysis	8	Schizophrenia	2015	558	US	Health comm/journalism and media
Zhang et al. (2014)	Framing, content analysis	N/S	Depression	2000-2012	1507	China	Journalism/media and culture
Zhang et al. (2016)	Content analysis	6	Depression	1980-2012	1656	US	Applied communication/ media & culture

TABLE 2.1. OVERVIEW OF EXISTING RESEARCH DISCUSSED IN SECTION 2.4 "PRESS DATA"

2.6 Existing research: an overview

As is evident from Table 2.1, the majority of studies to date that use newspaper data have been conducted in the field of psychiatry. Moreover, the existing research often uses content analysis or thematic analysis to discuss aspects of texts. As I have discussed at various points in this chapter, this means that the aspects of texts analysed are those that the analyst deems of interest, which may result in overly subjective/interpretative analyses. Furthermore, many of the studies use coding schemes which are only used in the field of psychiatry. Whilst such studies build on existing knowledge on mental illness reportage in psychiatry, they do not offer specific information about mental illness reportage to those working in other disciplines. For example, many of the coding structures used in the existing research do not go far beyond saying that an article is negative or positive, and relatedly whether it is stigmatising or not. The issue with research of this kind is that while is it invaluable in providing insights into trends in mental illness reportage over the years, it does not question what stigma is exactly, or how it is manifested in language (which is where researchers must believe stigma is manifested if newspaper articles are used as a unit of analysis). This is also true of the studies that use methods combining, for example, content analysis and statistical tests. They provide great insight into the trends within the data but neglect to recognise the wealth of information contained in each text that is more subtle that merely whether the text is positive or negative overall. My intention here is not to suggest that research of this kind is not useful. It is, and for the most part it is systematic. But while the methods employed are sufficient for meeting the research aims in psychiatry, there are inevitably analytical gaps that, from a linguistic standpoint, are glaring. It is clear that what is needed in research into the representations of mental illness in the news is for greater attention to be paid

to the systematic analysis of the linguistic structures in the text. This will offer a means of providing more nuanced understandings of mental illness. Here linguistics has a clear role to play. Very little research has been done by linguists in this area (aside from Atanasova et al., 2019), which is surprising given that (i) discourse analysis is a method familiar in the field and (ii) the linguistic analysis of mental illness more generally is a fertile area in linguistics (see for example, Demjen et al., 2019; Kinloch & Jaworska, 2019; Koteyko & Atanasova, 2018; Knapton, 2013; Harvey, 2012, 2014; Harvey & Brown, 2012; Hunt & Harvey, 2015; Tay, 2017). By way of an example of how linguistics can enrich findings in psychiatry, let us revisit the findings reported by numerous researchers that violence and criminality are key themes in news reports on mental illness. Research has shown that people with mental illness are viewed as being out of control of their actions, violent and dangerous. A simple collocation and concordance analysis can test whether there is a linguistic basis for this belief. I discuss the link between violence and criminality and mental illness in Chapter 7.

2.7 The place of the current study

As well as offering insight into some methodological limitations for researchers interested in language in mental health reportage, Table 2.1 also reveals some gaps in existing research. For example, there has been relatively little attention paid to the participants within the data other than their gender or age. In response to this gap in the research, Chapter 7 will analyse the naming strategies used to refer to people with mental illness within the data. The reason for doing this is that investigating how participants are named is a systematic and replicable way to explore how people with mental illness are discursively constructed in the press. This analysis then in turn gives insight into how

people with mental illness are viewed in UK society (because journalists write for the public).

Another gap in the research in UK-based studies is that studies tend to be small-scale. This is perhaps due to a tendency in the field to conduct qualitative data analysis. However, corpus linguistics offers methods to facilitate the detailed analysis of big data. In response to this gap in the research, this thesis presents analyses based on a corpus of over 50 million words, comprising articles published over a 30-year period. To my knowledge, this makes the current project the largest scale project carried out to date in this area. Moreover, the data used in this thesis is continuous over the 30-year period and does not rely on making diachronic observations of data using discrete datasets (e.g. Goulden et al., 2011; Roberts et al., 2013). Furthermore, as was noted in the previous section, previous research has found variation in the extent to which specific illnesses are stigmatised. The design of the data used in this thesis, which comprises illness subcorpora and year subcorpora, means that the idiosyncrasies of illness can be explored in detail.

2.8 Conclusion

In this chapter, I have shown that studies into the language of mental illness have spanned a variety of disciplines. In the existing research, researchers have made efforts to analyse how society views mental health, using mass media – and often newspaper discourse specifically – as data. This is based on the rationale that the press informs how society think about mental health (and arguably that newspaper discourse is a reification of current societal thought on mental health and therefore a worthy object of study). This idea that there is a link between texts and society is one echoed by Baker who writes "considering that corpora contain natural occurring data, they have the potential to tell us as much about the values of societies they came from as they

do about language" (Baker, 2010: 121). I have shown how this rationale is based on the theoretical position of social constructionism (Burr, 1995). To date, studies of mental health representations in the press that make use of linguistics are few; instead, the methods used are overly interpretative (e.g. close reading or content analysis) and applied to small datasets. However, there is an emerging body of research in linguistics into the language of mental illness (Demjen et al., 2019; Atanasova et al., 2019; Deamer & Hayward, 2018).

What unites existing studies in linguistics and studies from other disciplines into language is that language affects the way we see the world; i.e. that meaning is socially constructed. It is my contention that linguistics offers the tools to better understand the social constructs that are taken for granted in the existing literature on mental illness, constructs such as stigma. Moreover, the analysis of language provides a means of seeing social constructs emerge from the language, rather than looking for existing social constructs in the data. For this reason, a more nuanced approached to language analysis in this area offers new ways of looking at mental illness reportage.

There is, then, a clear gap in current research for studies that combine methods from linguistics and insights from mental health studies like those reported in this chapter. This thesis uses big data to analyse the discursive construction of mental illness with no preconceived expectations of the data and with no particular *a prioi* focus, such as the assumption of stigma or positive or negative portrayals of mental illness.

In this chapter I have described the theory underpinning this thesis, that of social constructionism. I have also reviewed the existing literature on stigma research and research into mental illness that uses language data. I have reported the key findings in the existing research in addition to identifying the methodological problems in the existing literature. Furthermore, I have positioned by own research in relation to the existing literature. In the next chapter, I describe the analytical methods I use from corpus linguistics.

3. Analytical Methods 1: Corpus Linguistics

In Chapter 2, I described the theoretical position underpinning the analysis in this thesis and reviewed the existing literature into mental illness and language. In this chapter, I detail the analytical methods and tools used in this thesis taken from corpus linguistics.

3.1. Introduction

In this chapter I provide a brief overview of corpus linguistics and the key debates and schools of thought in corpus linguistics. In addition, I describe each analytical method taken from corpus linguistics that I use in this thesis. Each section of this chapter will relate to a different method, e.g. keyword analysis, key semantic domain analysis. I exemplify each method using data collected during a pilot study of this thesis, and a sample of the MI 1984-2014 corpus. The two corpora used in this section are the *sample corpus* containing newspaper articles that discuss mental illness (559, 874 tokens) and a more specialised corpus containing newspaper articles containing the lemma *mania** (271,874 tokens), called the *mania corpus*.

In Section 3.2, I describe briefly the varying ways in which corpus linguistics is used, e.g. as a theory and as a method. In this section I set out the way that I use corpus methods in this thesis. In section 3.3, I describe the different corpus methods used in this thesis.

3.2. Corpus Linguistics

Corpus linguistics is the study of linguistic patterns in large quantities of machine-readable text "that defies analysis by hand or eye alone within any

reasonable timeframe" (McEnery & Hardie, 2012: 2)[12]. The texts under scrutiny may be written, or transcribed speech. The theoretical underpinnings of corpus linguistics can be traced back to the pioneers of contemporary linguistics and their work on language documentation, such as Franz Boas who famously wrote "While until about 1880 investigators confined themselves to the collection of vocabularies and brief grammatical notes, it has become more and more evident that large masses of texts are needed in order to elucidate the structure of languages" (Boas 1917: 1). In modern linguistics, corpus linguistic theory and methods are used in a wide range of subfields, such as those focused on the generation of grammars, natural language processing, and pedagogical linguistics in both English language and ESL contexts, right through to research such as that reported in this thesis that is concerned with analysing a particular variety of language within a fairly circumscribed context.

Broadly defined, there exist two schools of thought in corpus linguistics. The first is the neo-Firthian School which originated at the University of Birmingham and is rooted in the work of John Rupert Firth, the UK's first professor of linguistics[13] (see, for example, the work of such neo-Firthian scholars as Sinclair (1991), Hunston (2002), Louw (1993), Teubert (2005). Generally speaking, the neo-Firthian approach views corpus linguistics as a "sub-field in its own right" (Hardie & McEnery, 2010: 385) with its own theoretical status (Tognini-Bonelli, 2001). This type of corpus linguistics is characterised by the belief that the corpus should be the source of hypotheses about language, or as Teubert writes, "It is the discourse itself, and not a language-external taxonomy of linguistic entities, which will have to provide the categories and classifications that are needed to answer a given research question" (Teubert, 2005: 4).

[12] It is important to note here that early corpus linguistics was not necessarily computerised. For example. The Survey of English created at University College London was paper-based.
[13] J. R. Firth was made Professor of Linguistics at the University of London in 1944.

The second school of thought is rooted in the work of scholars such as Randolph Quirk at University College London and Geoffrey Leech at Lancaster University. In contrast to the Birmingham School, this approach views corpus linguistics as a method by which researchers can test hypotheses or intuitions about language. It is this approach that underpins much of the research aiming to explore ideology in language (e.g. Baker et al., 2008; Partington et al., 2013; Wright & Brookes, 2018; *inter alia*) and test intuitions about, for example, literary texts (e.g. McIntyre & Walker, 2019; O'Halloran, 2007; Short et al, 2002; Semino & Short, 2004; *inter alia*). These two schools of thought or 'traditions' (Hardie & McEnery, 2010) have been termed as 'corpus-driven' and 'corpus-based' respectively by Tognini-Bonelli (2001). In line with McEnery & Hardy (2012: 6), it is my view that these terms are not particularly useful because it is not the case that all corpus research neatly falls into one or the other 'camp'. Indeed the debate surrounding the corpus-based vs. corpus-driven distinction (that has come to be known as the 'bootcamp' debate)[14] has led proponents of both persuasions to make rather blunt observations of the other; for example Gries (2010: 330) claims that purely corpus-driven work is a "myth at best", and Teubert (2010: 356) writes that corpus linguistics "has been hijacked by theoretical linguistics of all feathers". The bootcamp debate led McEnery & Hardie to refer to the two traditions as the 'methodologist' tradition and 'neo-Firthian' tradition (Hardie & McEnery, 2010: 385). Hardie & McEnery (2010) state that they view themselves as methodologists because, as they point out elsewhere (McEnery & Hardie, 2012: 6), the binary view of corpus-based vs. corpus-driven can be discarded based on the fact that many researchers

[14] [14] The bootcamp debate gets its name from an email thread posted to the Corpora mailing list in which Stefan Gries referred to a corpus workshop as a 'bootcamp'. Subsequent replies to the thread took issue with the content of Gries' workshop, resulting in the discussion of what corpus linguistics was exactly, i.e. method or theory. The email thread came to be the basis of a special issue published in the *International Journal of Corpus Linguistics* (Pope, 2010).

working in the methodologist tradition reject "the notion that the corpus itself has a theoretical status, and thus also rejects the binary distinction between corpus-based and corpus-driven".

It is worth stating here that my own view is that corpus linguistics is a set of methods that I can use as a researcher as I see relevant to answer my research questions thoroughly and objectively. The reason I subscribe to this view is that fundamental to my interest in language is an interest in people and how people use language. As a result, a corpus can only help me investigate language usage so far. For example, corpus software facilitates the finding of all instances of the modal auxiliary verb *should* in a corpus, but what it cannot do is tell me whether each instance of *should* is deontic or epistemic, or what the effect was of using that verb on the participants discussed in the newspaper article. What corpus software does allow me to do is (i) triangulate my research by way of generating and testing hypotheses based on huge quantitates of data, (ii) test my intuitions about texts and (iii) provide me with objective parameters for the analysis of sub-corpora.

While I subscribe to the view that corpus methods should be used and discarded as is necessary, it is vital to be mindful of and explicitly state the underlying theory that one has to subscribe to when basing any conclusions on corpus data at any level, and particularly for thematic corpora. For example, if generating hypotheses from corpus findings, it is important to state that these hypotheses will be conditioned by the sampling frame used to compile the corpus. This is not a weakness of the method but a strength as it provides better representation of the variety being analysed. Additionally, it is important to draw attention to the fact that corpus linguistics is mostly concerned with the language at the level of the token and that this is at odds with other theories in linguistics, e.g. pragmatic theories. Even if you are only interested in language at the level of the word this entails a set of assumptions about language. It would be wrong to assume, therefore, that the methodologist approach is

entirely atheoretical. Basing conclusions on corpus evidence requires subscribing to very many underlying theoretical positions, the most obvious being that frequency analysis offers a means of exploring a word's significance. With this in mind, there is no reason why a linguist engaged in corpus analysis for the purposes of supplementing other analyses cannot develop the theoretical and methodological understanding of corpus linguistics any less than those engaged in research governed by the neo-Firthian tradition.

For all of these reasons, I will refer to the type of corpus analysis conducted here by yet another term, 'corpus-assisted' (cf. O'Halloran, 2007), which forms part of the bigger approach to the study of texts known as 'corpus-assisted discourse studies' or CADS (Partington, 2004, 2006). The reason for this is that this term better encapsulates what I view as my approach in this thesis, i.e. that corpus-assisted analysis uses methods from CL and CDA. In line with Baker et al. (2008: 274), it is my view that "neither CDA nor CL need be subservient to the other (as the word 'assisted' in CADS implies), but that each contributes equally and distinctly to a methodological synergy". Moreover, CADS is a well-used term in research with similar aims to mine (e.g. Baker et al, 2008; Partington et al., 2013; Wright & Brookes, 2018, *inter alia*).

In the following sections of this part of chapter 3, I will describe the methods used in the analysis section concerned with corpus linguistics. In what follows, I present each method separately, as doing so helps make clear the utility of the analysis and the statistical tests associated with the particular method. This is a somewhat simplistic way to present each method, as many need to be used in conjunction with others; for example, it is impossible to know the full significance of a particular collocation without concordance analysis. This is a view espoused by McEnery & Hardie (2012) who write that qualitative and quantitative analyses are "equally important to corpus linguists" (McEnery & Hardie, 2012: 3).

Moreover, the methods described here will be used in conjunction with methods taken from critical discourse analysis (CDA) in order to semi-automate the analysis of particular textual practices, e.g. modality. A full description of CDA methods is given in Chapter 4.

3.3. Corpus methods

In this section, I describe each corpus linguistic method used in this thesis. Each section will describe the method using data taken from the sample corpus and the mania corpus. I will also describe the statistical tests and statistical cut-offs associated with each method. In section 3.3.1 I describe frequency analysis, in Section 3.3.2-3.3.4 I describe methods associated with keyness analysis, specifically keyword analysis (Section 3.3.3) and key semantic domain analysis (Section 3.3.4). In Sections 3.3.5 and 3.3.6 I describe collocation analysis and N-gram analysis, respectively. Section 3.2.5 describes concordance analysis and in Section 3.2.6, I describe semantic preference and semantic prosody. In section 3.4, I conclude.

3.3.1. Frequency analysis

In its most basic form, frequency analysis is the measure of how often a word or phrase occurs within a corpus. Frequency analysis is often the first stage of corpus linguistic analysis, as it gives an indication of the topics discussed in a corpus. However, it is often the case that the highest frequency words in a corpus are function words that reveal very little about the themes within a corpus (this is not to say that function words do not reveal information about the corpus, as I will discuss in more detail in my discussion of style markers in Section 3.3.2. This means that the analyst may need to explore the lower ranking items on the frequency list in order to ascertain salient topics or words.

Frequency lists show all the types (distinct words) within a corpus grouped by frequency. Frequency lists are also known as wordlists.

In order to ensure that word frequencies are comparable across corpora of different sizes, frequencies are usually normalised to a common base by generating a relative frequency[15]. In the case of Wmatrix (Rayson, 2009), the software used to generate the wordlist shown in Table 3.1), the common base is 100. This means that the result will show how many times a word occurs per 100 words. Relative frequency is calculated as follows:

Relative frequency = (frequency of token ÷ total number of tokens) x base of normalisation

For example, to manually calculate the relative frequency of *the* in the sample corpus according the raw frequency reported in Table 3.1, we would carry out the following calculation:

(26926 ÷ 559,874) x 100 = 4.81 (to two decimal places)

Generating descriptive statistics, such as calculating relative frequencies, provides an overview of the data under scrutiny and allows for the comparison of smaller corpora with larger corpora and vice versa. [16] Table 3.1 shows a wordlist for the sample corpus with both raw and relative frequencies.

[15] Also called *normalised frequency*.

[16] It is worth noting that different corpus tools calculate tokens differently and, as a result, frequencies (and therefore relative frequencies) will vary across software. Before manually calculating relative frequencies, the conscientious student should make sure to note which corpus tool they collected the raw frequencies from in order to save themselves a headache of the mathematical kind. This advice comes from bitter experience.

Rank	Word	Freq.	Relative Freq.
1	the	26926	4.81
2	to	15025	2.68
3	of	14980	2.68
4	and	14156	2.53
5	a	13068	2.33
6	in	9639	2.72
7	that	7118	1.72
8	is	6796	1.27
9	it	5572	1.21
10	I	5211	1.00

TABLE 3.1. WORDLIST FOR SAMPLE CORPUS (TOP 10 WORDS BY FREQUENCY)

Depending on the software available and the utility of such analyses in answering specific research questions, frequency lists can also be generated for semantic domains or parts-of-speech (POS). One piece of software that automatically tags corpora for semantic category and POS is Wmatrix (Rayson, 2009). Using the in-built CLAWS[17] and USAS[18] taggers, Wmatrix can generate frequency lists that give information about the wider meaning being discussed in a corpus through semantic tagging (semtags), and grammatical information, through POS tagging. Table 3.2 shows the 21 top-level semantic categories in the USAS tagset.

17 Developed at Lancaster University, the Constituent Likelihood Automatic Word-tagging System or CLAWS is an automatic part-of-speech tagger for English and achieves 96-97% accuracy (see Garside, 1987).

18 The UCREL Semantic Analysis System or USAS tagger developed at Lancaster University is an automatic semantic tagger based on McArthur's *Longman Lexicon of Contemporary English* (McArthur 1981). The USAS tagger is 92% accurate.

A general and abstract terms	**B** the body and the individual	**C** arts and crafts	**E** emotion
F food and farming	**G** government and public	**H** architecture housing and the home	**I** money and commerce in industry
K entertainment, sports and games	**L** life and living things	**M** movement, location, travel and transport	**N** numbers and measurement
O substances, materials, objects and equipment	**P** education	**Q** language and communication	**S** social actions, states and processes
T time	**W** world and environment	**X** psychological actions, states and processes	**Y** science and technology
Z names and grammar			

TABLE 3.2. TOP-LEVEL SEMANTIC CATEGORIES IN THE USAS TAGSET

The 21 discourse fields can be further subdivided, e.g. semantic category B. THE BODY AND THE INDIVIDUAL[19] can be divided into B1. ANATOMY AND PHYSIOLOGY, B2. HEALTH AND DISEASE, B2+. HEALTHY, B2-. DISEASE, etc. Table 3.3 shows the top five key semantic domains in the sample corpus.

Semtag	Description of tag	Freq.	Relative Freq.
Z5	Grammatical bin	169020	30.19
Z8	Pronouns	50051	8.94
A3+	Existing	16471	2.94
Z99	Unmatched	11224	2.00
B2-	Disease	11068	1.98

TABLE 3.3. TOP 5 SEMANTIC CATEGORIES IN SAMPLE CORPUS

[19] In line with Wmatrix, and for clarity, I will use SMALL CAPS to indicate a semantic domain in the rest of this thesis.

Table 3.3 shows the top 5 semantic categories in the sample corpus. The value of Table 3.3 is that this allows us to gain insight into the language used in a corpus. It is also possible to see how the semantic categories and the highest frequency words found in the wordlist are similar. For example, the GRAMMATICAL BIN category contains the prepositions, conjunctions, etc. that constitute the most frequent words in the corpus shown in Table 3.1. Table 3.3 also shows some of the problems inherent in automatic semantic tagging. The 'Z99. UNMATCHED' category shows the lexical items that the software has not been able to successfully tag. Tagging errors may include lexical items that are acronyms such as *OCD*, neologisms that the software does not yet know, or words or phrases containing symbols that the software cannot process, such as website addresses or hyperlinks. In addition to the software being unable to tag particular lexical items, the mistagging of lexical items can also be a problem, especially when automatically tagging thematic corpora where lexical items may not be functioning according to their dictionary definitions but rather as part of a novel usage. For example, in a corpus of beer advertisements featuring heavily gendered language (Price, forthcoming), I found that Wmatrix tagged slang usages of the colloquial term 'tits' (referring to breasts) as 'LIVING CREATURES: ANIMALS, BIRDS, ETC'. This example is one where the mistagging was fairly obvious as this semantic category was unexpected in the context of beer advertisements; however, mistagging such as this may not be obvious in other corpora. As a result, it is necessary to manually check semantic tags as much as is feasible according to the project (and add specialised lexis to the Wmatrix dictionary).

Generating wordlists for corpora is also a necessary preparatory phase in order to conduct a keyness analysis. This type of analysis will be described in more detail in the next section.

3.3.2. Keyness analysis

Keyness analysis is the analysis of words whose frequency is statistically significantly higher or lower that would be expected when compared to a reference corpus. Scott (1997) writes:

> A key word may be defined as a *word which occurs with unusual frequency in a given text*. This does not mean high frequency but unusual frequency, by comparison with a reference corpus of some kind.
>
> (Scott 1997: 236, original emphasis)

Keywords can be positive (those words that occur more in one corpus than another), or negative (those words that occur less in one corpus than another). Keyness analysis gives an indication of the "aboutness" (Scott, 1999) of a corpus, or what is idiosyncratic about a corpus. Keyness analysis can be useful tool for gaining insight into and describing current or historic cultural trends or stereotypes within a text or collection of texts (Scott, 1997: 243) (notwithstanding the fact that there is not always a direct link between the use of specific lexis and cultural trends). This is clearly a valuable method for analysing potential societal change in diachronic thematic corpora. Furthermore, the comparison of one text with other (larger) texts is essential in defining what is distinctive about a particular text in terms of its 'style markers' (Enkvist, 1973: 25), or stylistic features of a text that can be seen as deviating from a norm particularly function words.

Keyword analysis is a well-established method for corpus comparison and has been part of the WordSmith software package since its first release (Scott, 1999, 2016). However, the analysis of POS and semantic keyness is a relatively new method made possible by Wmatrix (Rayson, 2009). Keyness analysis, or the comparison of one frequency list with another to calculate

statistically significant categories, can be carried out on frequency lists pertaining to the frequency of individual tokens, POS categories or semantic categories within a corpus, resulting in 'keywords' (derived from the comparison of word lists), 'key POS' (derived from the comparison of POS frequency lists), or 'key semantic domains'[20] (derived from the comparison of semtag lists).

Keyness analysis is a central aspect in many contemporary corpus linguistic studies, attracting much attention in edited volumes (Bondi & Scott, 2010; Archer, 2009) and analysed in a range of text-types ranging from literary texts to email communication; examples of keyness analyses include the analysis of keyness in Shakespeare's *Romeo and Juliet* (Culpeper, 2009), in comparisons of narrators' voices in a novel (Walker, 2010), in discourses of political correctness (Johnson et al., 2003), in NHS direct phone calls (Adolphs et al., 2004) and in emails sent by adolescents pertaining to their health (Harvey, 2013: 90). Keyness analysis is wide ranging as it affords analysts the opportunity to explore the semantic and grammatical properties of texts in much greater depth than is feasible without computational methods. Moreover, as Adolphs et al. (2004) point out, "keyword analysis also serves as a powerful hypothesis testing device and enables the analyst to cross-reference the results with his/her intuition about the transcripts." (Adolphs et al., 2004: 14).

3.3.3. Keyword analysis

Keyword analysis offers insight into what may be idiosyncratic about a target corpus (the corpus being analysed) compared with a reference corpus. Such analysis may be text internal, i.e. the comparison is being made between

[20] Sometimes referred to as 'key concepts'.

subcorpora created from the same original data, or text external, i.e. the comparison is being made with the text and another general corpus, e.g. the BNC. The most basic form of keyness analysis is the comparison of keywords. Table 3.4 shows the keywords in the mania corpus[21] compared with the sample corpus (here, our reference corpus), which contains articles discussing mental illness generally.

Item	01	%1	02	%2	LL	%DIFF
Mania	528	0.19	33	0.01+	955.92	3194.90
Bipolar	675	0.25	142	0.03+	876.27	878.90
Manic	427	0.16	32	0.01+	748.10	2647.90
I	4201	1.55	5211	0.93+	581.04	66.02
Manic_depression	343	0.13	72	0.01+	441.13	881.03

TABLE 3.4. POSITIVE KEYWORDS IN THE MANIA CORPUS COMPARED WITH SAMPLE CORPUS (CALCULATED USING WMATRIX (RAYSON, 2009)).

Table 3.4 shows the keywords (under the column header 'item'), the frequency of the keyword in the mania corpus (column header '01') and the percentage that this keyword occurs within the whole corpus ('%1') before detailing this information for the reference corpus (column headers '02' and '%2'). A '-' or '+' symbol denotes whether the keyword is over or underused in the target corpus. Wmatrix then lists the log-likelihood score (LL) and '%DIFF' which is an effect size measure, which is the size of the statistical difference between two variables.

Although the results in Table 3.4 may not be interpretatively surprising on the surface, i.e. we might expect *mania*, *manic* and *manic depression* to be key when compared with a general corpus of articles discussing mental illness, the keyword analysis does yield some results that can form the basis for further research questions. For example, why is the first-person pronoun overused and is this indicative that people are more willing to discuss their own experiences

[21] The mania corpus contains newspaper articles containing the lemma *mania**.

with bipolar disorder than with other mental illnesses? Moreover, overused keywords in specialised corpora give the analyst an insight into the words that are used to describe bipolar disorder, or that are used in relation to it. In order to fully explore the questions raised by a keyword analysis, it is necessary to look closer at the keywords in context.

Unlike calculating word frequencies using descriptive statistics like those carried out in Section 3.3.1, keyness is most commonly measured using a log-likelihood test, which is an inferential statistical test.[22] The log-likelihood test is used in corpus linguistics over other similar tests because it can account for data that does not have a standard normal curve (or normal distribution) (McIntyre & Walker, forthcoming: 131). Log-likelihood is essentially a test to work out to what extent a word occurs in a target corpus compared with a reference corpus more (or less) than would be expected by chance. Although corpus software can automatically calculate log-likelihood, it is important to know why log-likelihood is used and what steps lie behind a keyness statistic. In what follows of this section, I will explain log-likelihood using McIntyre & Walker (forthcoming) and Rayson & Garside (2000) as the foundation for what I report.

Log-likelihood is calculated by working out the total frequency of words in both the target and reference corpus minus the observed frequency of the word in question. Then, the observed frequency (which is the total number of times the word in question occurs in both the target and reference corpus) is multiplied by the total number of words in the target corpus and then divided by the total number of words in both the target and reference corpus in order to get the expected frequency for the target corpus (Rayson & Garside, 2000:

[22] For a detailed discussion of the validity of inferential statistics in corpus linguistics see Gries (2005) in response to Kilgarrif (2005). See also McIntyre and Walker (2019) for an explanation of null-hypothesis testing.

3)[23]. This process is then repeated for the reference corpus (this is how negative keyness is calculated). The last step in calculating keyness is to measure how far the expected frequency is from the observed frequency. This is the log-likelihood value. The more the observed frequency deviates from the expected frequency, the more confident we can be about how 'key' the item is in terms of its statistical significance. It is the log-likelihood value that "tells us whether the word (or item) whose potential keyness we have been calculating is indeed key, thereby avoiding the need for us to rely on subjective judgements about such matters." (McIntyre & Walker, forthcoming: 133). A log-likelihood score of 3.84 (equivalent to $p < 0.05$) means the analyst can be 95% sure that they have a significant result. All of the results reported in this thesis pertaining to keyness will have at a log-likelihood of at least 10.83 (equivalent to $p < 0.001$), meaning that the results reported have a 99.9% likelihood of significance. In addition to increasing confidence in results, setting a higher statistical cut-off is also a useful way to refine large corpora for detailed analysis.

3.3.4. Key semantic domain analysis

Once an analyst has a frequency list of the semantic domains for a target and reference corpus, key semantic domain analysis can be conducted by comparing the two lists. Table 3.5 and Table 3.6 show the key sematic domains in the mania corpus compared with the sample corpus.

[23] This information is available at http://ucrel.lancs.ac.uk/llwizard.html, along with a LL and effect size calculator.

Semtag	Description	01	%1	02	%2	LL	%DIFF
E4.1+	HAPPY	827	0.30	558	0.10 +	423.78	205.21
Z8	PRONOUNS	28132	10.35	50051	8.94 +	379.63	15.75
K2	MUSIC AND RELATED ACTIVITY	537	0.20	349	0.06 +	289.16	216.86
Q4.1	THE MEDIA: BOOKS	800	0.29	690	0.12 +	277.88	138.76
I2.2	BUSINESS: SELLING	660	0.24	560	0.10 +	236.25	142.70

TABLE 3.5. POSITIVE KEY SEMANTIC DOMAINS IN THE MANIA CORPUS COMPARED WITH SAMPLE CORPUS

Semtag	Description	01	%1	02	%2	LL	%DIFF
G3	Warfare, defence and the army; weapons	276	0.10	3659	0.65 -	1514.85	- 84.47
S8+	Helping	992	0.36	4800	0.86 -	714.08	-57.44
B2	Health and disease	253	0.09	1934	0.35 -	529.90	-73.06
E6-	Worry	586	0.22	3037	0.54 -	507.99	- 60.26
G2.1	Law and order	417	0.15	2165	0.39 -	363.18	- 60.34

TABLE 3.6. NEGATIVE KEY SEMANTIC DOMAINS IN THE MANIA CORPUS COMPARED WITH SAMPLE CORPUS

Whilst it is difficult to interpret these results without viewing the lexical items within the categories in context, it is possible to see some general semantic themes within the mania corpus; for example, the 'HAPPY' emotions which may be indicative of the mania aspect of bipolar disorder. Moreover, it is possible to see potential links between mania and some areas of public life indicated by the 'THE MEDIA: BOOKS' semantic category which subsumes lexical items such as '*writer, biography, memoir*'. Using key semantic categories as a starting point, we can hypothesise about certain possibilities, e.g. that mania or manic depression has been written about more than other mental illnesses.

It is also possible to use the initial results of a key semantic domain analysis to spot potential noise in the corpus. For example, MUSIC AND RELATED ACTIVITIES may be an unexpected category and therefore warrant further exploration. In this case, the overuse of this category is partly down to the search term *manic* returning articles pertaining to the UK rock band, the *Manic Street Preachers*. Once a source of noise in the corpus is ascertained, it is possible to account for that in further quantitative analyses by removing such instances.

3.3.5. Collocation analysis

Collocation is the name given to the process by which words co-occur statistically more significantly than would be expected by chance. For example, in British English, the lexical items 'fish' and 'chips' are collocates because they occur together frequently. We can test the collocational strength of two words or phrases by conducting statistical tests that show the statistical significance of words that co-occur. For example, we can test the collocational strength of the lexical items *fish* and *chips* in British English by searching for these words in the British National Corpus, or BNC, a collection of 100 million words of British English. We may wish to search particular lexical items in order to find out more about common phrases in a language variety and relatedly, find out about the culture of that language, or simply to test an intuition we may have about particular words.

In the BNC, 'chips' is the top collocate of 'fish', occurring in 14.81% of all instances of 'fish'. If we compare the collocates of *fish* in the BNC with its collocates in a corpus of American English, such as the Corpus of Contemporary American English (COCA), a corpus of 520 million words, we see 'fish' and 'chips' co-occurring in only 1.85% of instances of 'fish', then, can tell us something about the corpus being analysed by way of revealing set

phrases within the corpus; and depending on the data set, collocation analyses can allow analysts to generalise from the data being analysed to draw conclusions about a particular topic (e.g. common foodstuffs) or language variety (e.g. American English).

Collocation is typically analysed by searching for all the words that occur significantly within a 9-word window (referred to as a window of collocation) that includes the node word (the word being searched for) and the 4 words to the left and right of the node word, or L1-4 and R1-4. Table 3.7 shows the top five collocates of *mental** in a sample corpus of newspaper articles discussing mental illness calculated using AntConc (Anthony, 2018).

Rank	Freq.	Freq. (L)	Freq.(R)	Stat (MI)	Collocate
1	20	20	0	5.19695	Chief
2	11	11	0	4.18635	Link
3	11	11	0	3.89046	Stigma
4	15	15	0	3.14170	Combat
5	6	6	0	3.08447	MoD

TABLE 3.7. TOP FIVE COLLOCATES OF 'MENTAL*' IN SAMPLE CORPUS (SPAN L4-R4)

Taking each column in turn, from left to right, Table 3.7 shows the rank of the collocate in descending order, the total frequency, or how many times the collocate occurs in the corpus, how many times the word collocating with the searched word is found on the right or left of the node word, and the statistical significance of the collocation depending on the statistical test used. The final column shows the word that collocates with the node.

As is shown in Table 3.7, the top collocates of 'mental*' all occur to the left of the node word. The relatively small size of the corpus (and therefore the relatively low frequencies of the collocates) means that we cannot discern whether this is a pattern in the language generally, although this would be a

useful thing to do if we were interested in whether, for example, the phrase 'mental*' is more commonly pre or post-modified.

We can explore the collocation shown here in more detail by expanding the context of the collocation. Table 3.8 shows the collocations in context.

L4	L3	L2	L1	NODE	R1	R2	R3	R4
Chief	executive	of	the	**mental**	health	charity	Sane	welcomed
link	between	cannabis	and	**mental**	illness,	recent	studies	have
stigma	and	isolation	of	**mental**	illness.	his	prepared	contribution
combat	stress,	the	veterans'	**mental**	health	charity	which	is
MoD	spokesman	said:	the	**mental**	health	of	service	personnel

TABLE 3.8. TOP CONTEXT OF 'MENTAL*' IN THE SAMPLE CORPUS

Collocation is calculated using information about "the frequency of the node, the frequency of the collocates, and the frequency of the collocation" (Baker et al. 2008). There are a variety of statistical tests to measure collocational strength, however one of the standard statistical tests for collocational significance is Mutual Information (MI)[24]. Mutual information tests the dependence of two random variables, e.g. the lexical items 'fish' and 'chips'. The higher the MI score, the stronger the collocation. Whilst the validity of MI as a marker of collocation has been the subject of some discussion due to the fact that it can "unduly overvalue infrequent words" (Xiao & McEnery, 2006: 105), MI is the standard built-in statistical measure for collocation in much corpus software and, as Xiao & McEnery note, concerns about the validity of the measure can be partly alleviated by setting a higher minimum co-occurrence cut-off. (2006: 105).

McIntyre & Walker explain that if the MI score is higher than three, "then this can be taken as indicative of strong collocation" (Barnbrook, 1996; McIntyre & Walker, forthcoming; see also Hunston, 2001: 71-2). In contrast to this, Baker (2016: 142) argues in line with Durrant & Doherty (2010: 145) that to claim that a collocation is 'psychologically real'[25] "e.g. one word to trigger the thought of another, an MI of 6 would be required." (Baker, 2016: 142). My view on this is that a collocation need not be psychologically real in order for it to be interpretatively significant within the corpus, although of course a higher MI score increases the likelihood that the findings can be generalised from to make observations about the language as a whole. Moreover, the psychological reality of a word, e.g. whether one word is a predictor for another word, cannot

[24] Mutual Information is a common test in collocational analyses and is the default test in most corpus tools. Mutual Information is the only statistical test used for collocation analyses in this thesis.

[25] Durrant & Doherty use 'psychologically real' to refer to "psychological priming between words" (2010: 144).

fully be ascertained within most measures of collocation. As Gries (2013: 141) argues, collocational strength is typically calculated based on a bidirectional statistical test, which means that the direction of collocation cannot be ascertained.

As the corpora under scrutiny in this thesis are thematic, diachronic and representative of a particular variety, i.e. newspaper discourse discussing mental illness, the aim is not to generalise about language use beyond the specialist population from which the corpus is sampled. Moreover, in reference to Baker's claim that collocates should have an MI score of 6 or above, the problem with limiting collocates to particular statistical cut-offs is that it limits the ability to trace changes in collocation through time which I see as one of the key aims of diachronic analysis. For this reason, the minimum MI score used in this thesis will be 3.

As well as telling the analyst something about a corpus more generally, for example features of the topic or variety, collocation is also useful for revealing information about the meaning of specific lexical items or units of meaning, how they are used and how they interact with other words. Studying meaning through collocation, particularly in the field of distributional semantics, was pioneered by J. R Firth who famously stated "you shall know a word by the company it keeps" (Firth, 1957: 11). The Firthian view of meaning encompasses two important features:

(a) (*in observation*) that meaning and context are inextricably linked and we cannot analyse meaning without also taking context into consideration *viz* words that commonly occur in similar contexts have related meanings,

and,

(b) (*in usage*) that if meanings are related, then there must be a psychological basis for collocation

The Firthian view of collocation being indicative of word meaning has come to form the bedrock of many contemporary 'neo-Firthian' analyses (Hunston, 2002; Louw, 1993; Stubbs, 2001). Firth's original premise also forms the basis of semantic prosody (see section 3.2.6) and the theory that underpins the idiom principle which states that text production is made up of "semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments" (Sinclair, 1991: 110). It is also this view that forms the basis of much corpus-driven linguistic research into units of meaning, or broadening the unit of analysis in corpus linguistics from word to multi-word phrases (see Danielsson, 2013).

In diachronic corpora, collocation analysis can offer insight into lexical and phrasal changes in corpora that focus on particular topics. For example, in an analysis of the representation of refugees and asylum seekers in the UK press, Baker et al. (2008) posit two methods for dealing with changes in collocates over time: determining 'seasonal collocates', which they define as "collocates that are very frequent in a small number of years" (Baker et al, 2008: 286), and 'consistent collocates' (also known as c-collocates), which are present in most of the sub-corpora. Given the underlying assumption made in studies that use corpus linguistic analysis to analyse societal issues, analysing changes in collocation along with consistent collocates in diachronic corpora offer an insight into how a concept has changed or stayed the same over time.

3.3.6. N-grams

Related to the concept of collocation is the concept of n-grams[26]. N-grams are strings of lexical items or characters that occur frequently within a corpus (*n* stands for any number). Due to the fact that this thesis will only discuss n-grams at the level of the word, all future reference to n-grams will relate to lexical n-grams. N-grams are typically studied by their frequency rather than their statistical significance given that, unlike collocation, n-grams are not tied to a node word (a two-word n-gram is known as a bigram).

N-grams are well-used in the field of natural language processing where lexical strings are used as a tool to model and predict, amongst other things, the syntax of a given language, or the next word or character within a given lexical string[27]. In linguistics more specifically, n-gram analysis has been applied in the field of subtitling (McIntyre et al. 2018), in research into formulaic sequences (Buerki, 2016), and in the exploration of identifying authors based on their language use, e.g. in research into plagiarism (Johnson & Woolls, 2009: 112, Johnson, 2013), authorship attribution (Coulthard & Johnson, 2010, Grieve, 2007; Wright, 2013) and stylometry (Hoover, 2001). Such research provides empirical evidence for the psychological basis for language clusters which is a useful starting point for the linguist exploring whether language use can reveal societal thought on a particular topic, e.g. mental health. In addition to this, n-grams are useful for simply giving an indication of common words and/or phrases within texts.

Table 3.9 shows the top five 4-grams in the sample corpus:

[26] Sometimes called 'clusters' or 'lexical bundles'.
[27] It is this type of research that underpins the predictive text systems on smart phones, computers, etc.

Rank	Freq.	Range	Cluster
1	386	1	Post traumatic stress disorder
2	126	1	Traumatic stress disorder ptsd
3	101	1	Per cent of the
4	95	1	With mental health problems
5	90	1	Obsessive compulsive disorder ocd

TABLE 3.9. TOP FIVE 4-GRAMS IN THE SAMPLE CORPUS

Table 3.9 details the rank, frequency, range (how many of the uploaded corpora the n-gram appears in) and the string. What can be seen in this example is that most of the top n-grams are the names for mental illnesses. This is interesting when we consider what this means in the context of labelling mental illnesses. For example, the very fact that post-traumatic stress disorder is listed is indicative of the recognition of it as a mental illness. PTSD was only recognised as a mental illness in the 1980s and PTSD symptoms had been known previously as 'shell shock' or 'combat fatigue'. In line with this, the cluster 'with mental health problems' is also indicative of a change in the nomenclature of mental illness as 'post modifying' or person first language is now the preferred means of describing people with mental illnesses, i.e. *person with schizophrenia* in contrast to *a schizophrenic*. Further to this, cluster or n-gram analysis can give insight into potential data skew. For instance, PTSD makes up around 64% of the total cluster tokens in the sample corpus which suggests that many more articles have PTSD as their focus than other mental illnesses. In this respect, n-grams also have a methodological role to play in the process of determining representativeness and balance in corpus construction.

3.3.7. Concordance analysis

Concordance analysis is the point at which quantitative and qualitative methods in corpus linguistics meet. Most commonly, concordancing involves

looking at lexical items in their sentential context (or in the context of the utterance if analysing spoken discourse). This involves the close analysis of the word or phrase of interest as it appears in the corpus, usually with several words either side. However, concordancing need not be focussed at the sentence or utterance level. McEnery & Hardie (2012), for example, state that concordances may range from the analysis of suffixes to multi-word expressions (McEnery & Hardie, 2012: 35).

Concordance analysis is a vital step in analysing how a specific word or phrase functions in a sentence or utterance in terms of its grammatical properties, but also what a word means within a given context.

In addition to revealing previously undiscovered meaning within a corpus, looking at concordance lines is one way that an analyst can check the accuracy of automatic taggers. For instance, without concordance analysis, the earlier example given in section 3.2.1, in which Wmatrix tagged the lexical item 'tits' as belonging to the LIVING CREATURES: ANIMALS, BIRDS, ETC. semantic category, may have gone unnoticed and, with it, a potential line of enquiry into the gendered nature of the advertisements being analysed. Furthermore, a bird's eye view of data may not properly convey the meaning of a word in context. For example, based on the frequency of the derogatory words 'psycho' and 'schizo' in the sample corpus it might be sensible to assume that a discourse of stigmatisation of mental illness exists; however, on further inspection many instances of these words appear in articles featuring meta-linguistic discussion about correct and incorrect terms for mental illnesses.

A core endeavour of corpus linguistics is the uncovering of patterns across instances of language in use with the aim of discovering grammatical patterns within a variety, or gaining insight into a how a particular topic is discussed. A key principle of the study of language *in use* generally, but particularly in the corpus analysis that accompanies discourse analysis, is the recognition that meaning is co-constructed and that language users negotiate

meaning by 'languaging' (Halliday, 1985a). It is this presumption of (minimally) a message, a sender and a receiver that has to underlie any notion of ideology in texts. Therefore, it is vital that researchers using corpus methods recognise the equal value of qualitative and quantitative methods in analysing texts. This is a sentiment shared by Stubbs who emphasises 'the need to combine the analysis of large-scale patterns across long texts with the detailed study of concordance lines' (Stubbs, 1994: 212). Table 3.10 shows a concordance table from the mania corpus.

in which a patient has alternating	**moods**	of mania, or hypomania, and depression
rollercoaster journey through extreme	**moods**	is devastating to spouses, families
in which the victim has alternating	**moods**	of manic over-excitement and of depression
was used to describe people whose	**moods**	swung from elation to despair
bipolar disorder is not curable, but the	**moods**	can be controlled, and even prevented

TABLE 3.10. CONCORDANCE OF THE LEXICAL ITEM 'MOODS'

Table 3.10 shows the many ways in which 'moods' is used within the mania corpus and how it functions in sentences designed to describe periods of mania (characteristic of bipolar disorder, previously known as manic depression). Although we only have access to five concordance lines for this example, it is already possible to see features of the language that may not come to light in other corpus linguistic analyses that are limited to the word level. One example of this is metaphorical language, such as the ROLLERCOASTER metaphors used to describe the alternating moods characteristic of mania. Discovering a metaphor such as this one opens up a new line of enquiry in the analysis. For example, whether it is the case that the ROLLERCOASTER metaphor is commonly used in reference to mania and whether this metaphor is only used in reference

to mania. An analysis of metaphor falls outside the scope of the analysis in this thesis, but such examples show the utility of concordance analysis. This is because research questions such as the ROLLERCOASTER metaphor example are only possible through close reading of the corpus, as linguistic devices like metaphor, which may have no set linguistic form, may not be revealed through corpus analyses. Once it is known that mania is described as a journey on a rollercoaster, the lexical items searched for within the corpus can be refined in order to reveal potential instantiations of a specific metaphor (e.g. 'ride', 'journey', 'ups', 'downs', 'turns'); however, metaphors cannot be found reliably automatically despite much work in this area in NLP[28] (see for example, the MetaNet project[29]) and some work in linguistics (see for example, the use of the USAS tagger in metaphor identification (Koller et al., 2008; Semino et al., 2009)[30].

Given that perhaps the main affordance of conducting corpus linguistic analysis is that it allows analysts to analyse huge quantities of texts, it is unfeasible to analyse every concordance line (and indeed to do so might be unnecessary for the research question being investigated). In order to analyse concordances in a more rigorous way, corpus tools have been refined to deal with more specific search queries. Such tools are referred to by McEnery & Hardie (2012: 43-46) as 'fourth-generation' concordancers'. AntConc (Anthony, 2018) and Sketch Engine (Kilgariff et al., 2004), for instance, are optimised for regular expressions, while other tools such as CLIC (Mahlberg et al., 2016) even

[28] Refining methods for the automatic identification of metaphor has been a pursuit of the Natural Language Processing community for some time now; however, the basis for machine recognition of metaphor is still reliant on a human input, e.g. eliciting metaphoric set expressions from participants. Even if successful, automatic metaphor identification is limited in how many metaphors can be identified given that some novel metaphors may never have been uttered before, meaning that the software does not have the parameters needed to identify it.

[29] https://metanet.icsi.berkeley.edu/metanet/

[30] http://www.research.lancs.ac.uk/portal/en/upmprojects/using-a-semantic-annotation-tool-for-research-on-metaphor-in-discourse(af625fbc-b8d6-4972-b571-f35ade5e69e0).html

have in-built functionality for filtering concordances by subcorpus or search term. Despite the ability to refine concordance results, corpus software is still fairly limited in affording the analysis of linguistic patterns that fall between grammar and semantics such as transitivity processes (Halliday, 1985b; see Chapter 4). The first reason for this is that the software it is not yet sophisticated enough to code concordances for, e.g. agent, action, patient relationships, that require manual annotation. The second reason is a practical one: textual analysis requires a consideration of cohesion and viewing a chunk of text within a software window is not conducive to this.

As a result of this, much research combining corpus linguistics and (critical) discourse analysis makes use of downsized samples of corpora in order to subject them to more in-depth qualitative analysis, e.g. for euphemistic language or to identify types of modality. Recent research has explored the benefits of combining methods to respond to criticisms of cherry-picking examples in CDA and to exemplify the utility of mixed methods for research triangulation (Baker & Levon, 2015) and to incorporate methods from corpus linguistics into the 'academic movement' of CDA (Baker et al., 2008: 274).

3.3.8. Semantic preference and semantic prosody

In this final section on corpus methods, I describe the related notions of semantic preference and semantic prosody. This section provides a natural segue into the more qualitative methods described in Chapter 4 on methods from critical discourse analysis (CDA), as both semantic preference and semantic prosody require the interpretation of concordances and expanded contexts.

In Section 3.3.5 on collocation, the notion of semantic preference was introduced in reference to neo-Firthian research. An example of this is Sinclair's (1991) work on the idiom principle that states that texts are composed of set

phrases or semi-preconstructed phrases rather than discrete words with their own individual definitions, e.g. idiomatic expressions, proverbs, phrasal verbs, etc. (Sinclair, 1991: 110). The argument for the phrasal basis for language is a convincing one, evidenced by research that has observed recurring sequences in concordances (Sinclair, 20001, 2004) and convincing psycholinguistic evidence that has shown that people with speech and language disorders such as aphasias preserve the ability to produce formulaic language (Lum & Ellis, 1994; Wray & Perkins, 2000). Moreover Wray (2002) reports that speakers pause less when producing formulaic utterances. This compelling evidence suggests that "formulaic sequence meaning has cognitive reality" (O'Halloran, 2007). What unites the notions of semantic preference and semantic prosody with Sinclair's idiom principle is that they are all concerned with collocation in language and that they are a product of the 'phraseological tradition' (Hunston, 2007).

Due to the fact that semantic preference and semantic prosody are both "a collocational phenomenon and one which is preferably to be regarded as recoverable computationally from large language corpora rather than intuitively" (Louw, 2000: 48), they have been conflated in some previous research (see (Bednarek, 2008; Hunston, 2007 for a discussion). To avoid the problems associate with this, I will focus first on semantic preference and then on semantic prosody, but for the sake of clarity it is worth pointing out at this early stage that semantic preference is a property of individual lexical items, whereas semantic prosody is a property of units of meaning (McIntyre, 2018).

The semantic preference of a lexical item can be described as the type of semantic context it is commonly found in. Semantic preference is defined by Sinclair as "the restriction of regular co-occurrence to items which share a semantic feature, for example that they are all about, say, sport or suffering" (Sinclair 2004: 142). By way of an example, Stubbs (2001) looks at collocates of word phrase *ripe*. He found that *ripe* was often used in the phrase *ripe for* and

was found in contexts discussing change, and particularly change in reference to the climate, conditions, situation or time. Stubbs also found that when a noun phrases followed *ripe for*, it often had negative connotations (2001: 456-457). The interpretation of a set of collocates as being typically positive or negative through concordancing is the foundation of the notion of semantic preference whereby analysts look for robust corpus evidence to suggest that a particular word occurs with negative/positive collocates. Bednarek (2008) makes the point that whether a set of words are viewed as positive or negative is ultimately down to the analyst's subjective opinion, and that "semantic preference is probably context-, genre- and domain dependent" (2008: 123). Bednarek's point here is that whether a word routinely collocates with positive or negative words will depend on the context of the text under examination and therefore stating that a word has a set negative or positive semantic preference is problematic.[31]

Semantic prosody is concerned with the discourse function of a unit of meaning (Hunston, 2007). Louw (1993: 157) defines semantic prosody as the "consistent aura of meaning with which a form is imbued by its collocates" [32]. In a departure from Louw's original conception of semantic preference, some researchers have used semantic prosody to refer to whether the collocates of a word are positive or negative (e.g. Partington, 2004) which has caused confusion as to how semantic prosody and preference differ. By way of a response to this confusion, Hunston (2007: 266) argues that semantic preference should be referred to as 'attitudinal preference' and semantic prosody should

[31] It is worth saying here that while I agree with Bednarek's point as a general rule, I do not think this is such an issue in thematic corpora where the analyst's aim is to look at how words/units of meaning function within a specific variety. The possible conclusions that can be derived have limited generalizability by the very nature of the corpus being analysed anyway.

[32] Although Louw was the first to use the term *semantic prosody* in print, he attributes it first to Sinclair (see Louw, 2000: introduction).

be restricted to the definition provided by Sinclair (2004) which relates only to the discourse function of a unit of meaning.

For the purposes of this thesis, and to return to the criticisms made by Bednarek (2008) (see also Whitsitt, 2005), I take the view that the semantic prosody of a unit of meaning and the semantic preference of a word is more nuanced than simply being positive or negative, which is what Partington (2004) suggests.

3.4. Conclusion

In this this chapter I have provided a brief overview of corpus linguistics and its status as a theory and as a method. Furthermore, I detailed the view of corpus linguistics that I take in this thesis where corpus linguistics is used as a method. I also provided an overview of the basic corpus methods and analyses used in the analysis chapters of this thesis. In Sections 3.3.1-3.3.7 which each explored a different method, I also provided a discussion about the utility of each method enriched by data from two corpora containing newspaper articles reporting on mental illness.

In the next chapter 'Analytical Methods 2: Critical Discourse Analysis, I provide an overview of CDA and the analyses from CDA that I use in this thesis.

4. Analytical Methods 2: Critical Discourse Analysis

In Chapter 3, I described the analytical methods used in this thesis from corpus linguistics. In this chapter I provide an overview of the analytical methods used from critical discourse analysis.

Section 4.1 will provide an overview of CDA covering the early manifestations of linguistic inquiry into ideology in texts such as that of the East Anglia School (Fowler et al., 1979) to contemporary research into corpus-assisted discourse analysis that combines these early principles of CDA with computational methods. In Section 4.2, I describe the analytical methods used in this thesis in turn. In Section 4.3, I conclude.

4.1. Critical Discourse Analysis

Critical Discourse Analysis (CDA) can be described as an interdisciplinary approach to text analysis, which aims to uncover covert meanings in texts, usually with a view to uncovering power asymmetry or manipulation strategies. Whilst the remit of CDA does not preclude the study of non-linguistic features of the social artefacts under scrutiny (say, the visual aspects of a text), as a linguist working under the umbrella of CDA, I see my role as dealing with the linguistic aspects of texts. For this reason, this section will describe the linguistic analytical methods associated with uncovering meaning in texts.

CDA can be traced back to the East Anglia School that developed in the 1970s and the work conducted there into Critical Linguistics. Key figures in this movement were Roger Fowler, Bob Hodge, Gunther Kress and Tony Trew, whose work on the book *Language and Control* (1979), formed the basis of much of the contemporary work in CDA (see also Kress & Hodge, 1979); that is, the analysis of the way that "language functions in social and political practice"

(Fowler et al., 1979:1). Like much of the CDA that followed[33], Fowler et al. (1979) borrowed the functionalist theory from Halliday's (1978) systemic functional linguistics (hereafter SFL) which

> grew out of an effort to develop an **applicable** kind of linguistics […] drawing on functional and anthropological approaches to language in Europe and North America from the 1920s […] where theory is designed to have the *potential* to be applied to solve problems that arise in communities around the world, involving both reflection and action.
>
> (Matthiessen 2012: 436, original emphasis)

Both theorists of SFL and Critical Linguists posit that "language is as it is because of its function in social structure" (Halliday, 1973: 65). As such, both researchers working in SFL like Halliday, and those working in Critical Linguistics like Fowler, Kress and Hodge, are interested in the relationship between the form and the function of language and how this can create ideological meaning. Fowler & Kress write of the link between form and function, "we follow Halliday in requiring that social meanings and their textual realizations be included within the scope of a grammatical description" (1979: 187). This is based on the assumption that "the structure of a language should generally be seen as having been formed in response to the structure of the society that uses it" (Fowler & Kress, 188). The study of the form and function of language with a view to be 'socially accountable' (Mattheissen, 2012) has led researchers adopting SFL to refer to their work as 'applicable

[33] It is only very recently that researchers have started to propose different methodological perspectives for the analysis of ideology in texts. One methodology gaining particular traction is Cognitive Linguistic Critical Discourse Studies (Hart & Lukeš, 2007, Hart, 2017) which researchers argue provides "something like a conceptual account of Halliday's ideational function of language" (Hart & Lukeš, 2007: xi).

linguistics' which is a synthesis of the thesis/antithesis[34] positions in linguistics where researchers are either *theoretical* or *applied* (Mattheissen, 2012: 437).

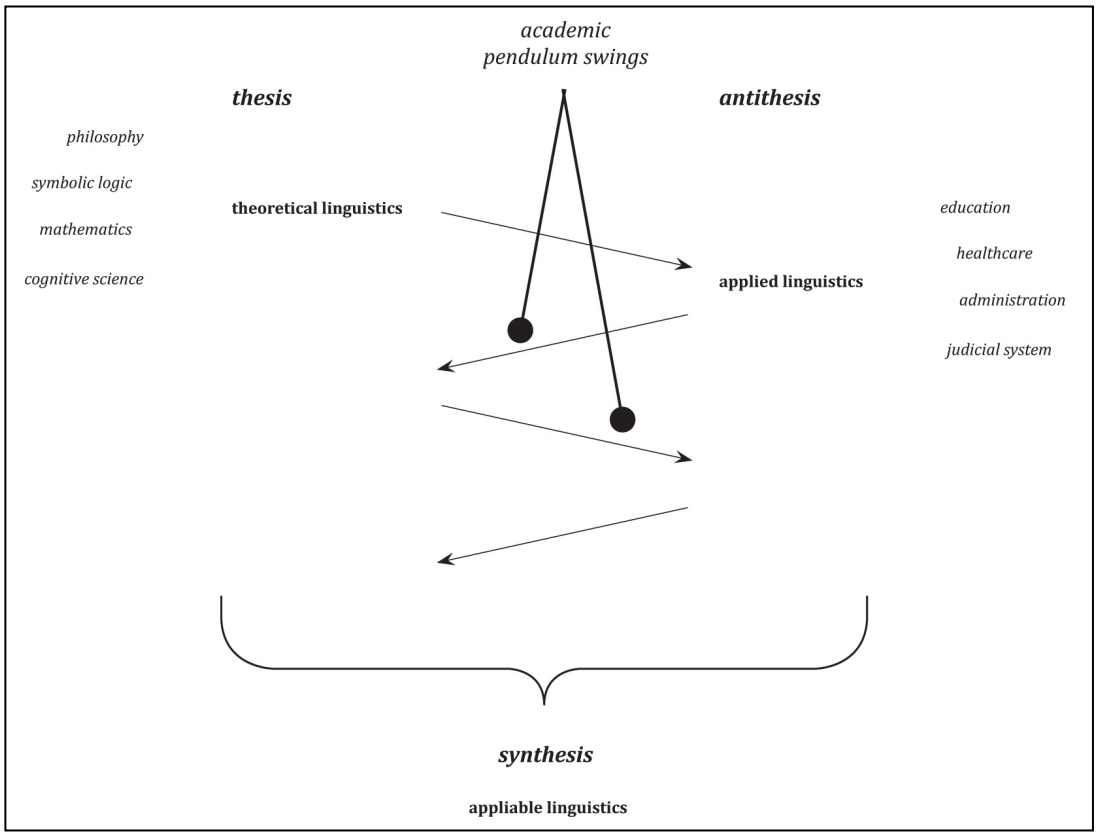


FIGURE 4.1. APPLICABLE LINGUISTICS AS A SYNTHESIS (TAKEN FROM MATTHEISSEN, 2012: 437)

The Hallidayan viewpoint that Critical Linguistics borrows (i.e. that researchers can conduct linguistic analyses with a view to being 'socially accountable') is clearly based on the assumptions that underlie linguistic relativity. It is clear in Halliday's writings that the principles of SFL are guided by the Sapir-Whorf Hypothesis and later, the works of Malinowski (1923), who saw language as a "mode of action" (Malinowski, 1923: 312). This in turn influenced Firth's (1957) view that context contributes to meaning and that

[34] Thesis and antithesis refers here to two contrasting views about what constitutes linguistic inquiry.

language has a functional basis (i.e. to be meaningful *to the participants in a linguistic event*) (see Bateman, 2017: 14 for a discussion of Firth's contribution to SFL). Context is crucial for Halliday who writes "language does not passively reflect reality; language actively creates reality. It is the grammar plus the vocabulary […] that shapes experience and transforms our perceptions into meanings" (Halliday, 2003 [1990]: 145). Put simply, for Halliday (and for the necessary underlying assumptions in Critical Linguistics) "language is a system for meaning making" (Halliday, 1985: xvii). In his SFL framework, Halliday posited three metafunctions (see Table 4.1 below) or "component[s] of meaning" (Halliday, 2003 [1973]: 314) that make up the "basic architecture of human language" at the lexicogrammatical level (Halliday, 2003: 16).

	Ideational	**Interpersonal**	**Textual**
METAFUNCTION	"clause as process" (Halliday, 2003 [1973]: 315)	*"clause as speech act"* (Halliday, 2003 [1973]: 315)	*"clause as message"* (Halliday, 2003 [1973]: 315)
DESCRIPTION	How language creates (and mediates) reality. The ideational metafunction is further categorised: *Experiential* How language shapes our perception of the world by "patterns of meaning that are installed in the brain" (Halliday, 2003: 15) *Logical* How language "set[s] up logical-semantic relationships between one clausal unit and other" (Halliday, 2003: 17). This process is iterative unlike the experiential system which is configurational. *Associated analyses:* Transitivity analysis, labelling, agency	How language creates and maintains relationships. How language is used to communicate "expressions of attitude and appraisal" (Halliday, 2003: 16). *Associated analyses:* Modality, sentence mood	How language is organised to convey a message. *Associated analyses:* Information structure, theme and rheme

TABLE 4.1 HALLIDAYAN METAFUNCTIONS

It is necessary to present the metafunctions as separate components; however, Halliday makes the point that the "structures deriving from the […] metafunctions are mapped on to each other" (Halliday, 2003: 18). A key feature of the Hallidayan viewpoint is that meaning is tied closely to language use. In fact, he labels language "the most complex semiotic system we [humans] have" (Halliday, 2003: 2). Crucially for CDA, Halliday's metafunctions, and in particular the ideational metafunction, identify the top-down nature of meaning as mediated by language (the experiential function), and the bottom-up construction of meaning (the logical function). For this reason, it is unsurprising that linguists interested in how language shapes human understanding of the world and our relationships within it, the *critical* linguists, adopted this model and in particular the view that language has an ideational function. Writing of Hallidayan linguistics in Critical Linguistics, Fowler states that the benefits of the three functions are that the formal features of language are

> conceived of functionally: not merely as formally different kinds of structure, but as kinds of structure that are as they are because they do particular jobs. […] The functions also provide a facility […] a prediction by theory of what types of linguistic construction will be partially revealing for critical linguistics. It is quite clear that the ideational and interpersonal functions are especially valuable […] since critical linguistics is particularly concerned with the ordering of experience and with the mediation of social relationships and values.
>
> (Fowler 1991: 70)

Fowler's work in the area of Critical Linguistics was influenced by the principles underlying Hallidayan SFL, and although contemporary research in CDA can be traced back to the work of the East Anglian School, the term

Critical Discourse Analysis was coined by another key figure in the analysis of ideology in texts, Norman Fairclough.

Fairclough's CDA, unlike that of the East Anglia School, is not primarily concerned with linguistic instantiations of ideology in texts. In fact, Fairclough argues that "there tends to be too much emphasis upon the text as product, and too little emphasis upon the processes of producing and interpreting texts" (Fairclough, 1992: 27). Rather, for Fairclough CDA is transdisciplinary (i.e. not necessarily linguistic) and as such, 'discourse', which Fairclough defines as "a mode of action, one form in which people may act upon the world and especially upon each other, as well as a mode of representation" (1992: 63) is analysed through looking at the semiotics (i.e. any meaningful element – not just text) of social life (Fairclough et al, 2004). For Fairclough, CDA is the study of discourse in this specific sense, through the analysis of semiosis – or networks of meaning - in the social world. Using this definition of CDA, Fairclough's research has focussed on the macro aspects of social life, or those that may not have a clear relationship with a linguistic form or device such as the process of naturalization (Fairclough, 1989: 76). Naturalization is the process whereby an ideological viewpoint that may be initially foregrounded becomes accepted as common sense (even though it may not be) through a variety of semiotic modes, and as a result, the fact that it remains an ideological attitude is backgrounded. It is fair to say, then, that Fairclough's research is concerned with the macro aspects of social life, which may not be easily explained by the kind of form/function analysis posited by the Critical Linguistics of the East Anglia School. Fairclough's focus on macro aspects of social life such as naturalisation has received criticism precisely because of the distance between such sociological descriptions and linguistic form. For example, Stubbs (1997), writing about naturalization, states "CDA presents no theory about the role of repetition in such influences. In common with

linguistics in general, it has no theory of how our ways of seeing the world are influenced cumulatively by repeated phrasings in texts" (Stubbs, 1997: 6).

Fairclough's notion of CDA has been widely adopted within and outside of linguistics, for example in politics (Dillon et al., 1993) and education (Rogers, 2004). As a combined result of non-linguists adopting methods from CDA, and increased attention paid to the macro aspects of social life without attention paid to language, some critics have argued that contemporary CDA has moved too far away from text analysis, and instead focuses too greatly on power on the "contextual (and thus necessarily somewhat vague) features of powerful language" (Jeffries, 2010: 1). It is worth stating however, that for some prominent critical discourse analysts, such as van Dijk (1995), CDA was never a field that belonged to a particular subdiscipline (1995: 17). Fairclough states that CDA "is not just the analysis of discourse (or more concretely texts), it is part of some form of systematic transdisciplinary analysis of relations between discourse and other elements of the social process" (Fairclough, 2010 [1995]: 10)[35]. Further to this, van Dijk writes that CDA is "problem- or issue-oriented, rather than paradigm-oriented. Any theoretical and methodological approach is appropriate as long as it is able to effectively study relevant social problems, such as those of sexism, racism, colonialism and other forms of social inequality" (1995: 17). It is worth stating at this point that van Dijk's definition of CDA differs from that of the Critical Linguistics outlined by the East Anglia School, for whom the analysis of social factors was a linguistic endeavour and stemmed from an *intra*lingual exposition of the thesis of Edward Sapir and Benjamin Lee Whorf (Sapir, 1921; Whorf, 1940). Van Dijk's conception of CDA gives rise to two issues. The first is that the object of study in CDA is language (and therefore it seems sensible to suggest that linguistic methods be used –

[35] Interestingly, Fairclough makes this suggestion that CDA need not be the analysis of texts in his 2010 [1995] monograph *Critical Discourse Analysis: The Critical Study of Language* (my emphasis).

which then negates van Dijk's suggestion that CDA does not conform to a particular paradigm). The second related issue is that if we take van Dijk's aims of CDA literally (i.e. that we have a goal [to study social problems, namely finding racism or sexism, etc. in texts] and an object of study [language] but no concrete definition of the approach [or in fact what constitutes that approach in terms of a set of tools]), then what we are left with is too vague to be usable and too goal-oriented to be objective. Arguably, research conducted according to these aims cannot purport to abide by the scientific principles that it ought to. Moreover, an approach with no set methods or theory but *a priori* aims (i.e. to find and analyse racist language) cannot generate research that is replicable, objective and rigorous. Critics of CDA have drawn attention to these issues. For example, Jeffries (2010) responds to the lack of a formalised linguistic framework in CDA by returning to the linguistically-driven critical analysis conducted by Fowler et al. (1979). Jeffries (2010) references Fairclough's three stages of CDA, "description, interpretation and explanation" (Fairclough, 1989: 26), to state that her impression of "the CDA literature generally" (2010: 11) is that researchers in CDA are focussed on the explanation stage rather than the linguistic description stage. As a response to this, Jeffries' (2010) 'Critical Stylistics' framework uses a set of linguistic tools (types of analysis) based on general grammatical analysis, but applied to the analysis of ideology in texts. Such analyses include labelling, nominalization, and transitivity under other names, e.g. naming and describing, representing actions/events/states. The utility of the Critical Stylistics toolkit is not in developing new tools but in formalising established ones and explicating their functions more explicitly for the textual analysis of power and ideology in texts.

Critical Stylistics is a useful toolkit for resolving the problems I previously detailed with van Dijk's conception of CDA, where its primarily goal is to look for racism, sexism, etc. without letting linguistic description uncover these ideologies in texts. I don't believe it is the case that CDA generally does not

focus on linguistic description; there are very many studies in CDA that do this clearly. For example, the following articles use nominalization (Dunmire, 2005; Flowerdew, 2002), transitivity analysis (Dreyfus, 2017; Fairclough, 1995; Seo, 2013) and labelling (or naming) (Jones et al., 2017, Page, 2003). However, I do believe it to be the case that there is a lack of a systematic linguistic framework for text analysis in CDA which can result in the tendency to rely on the explanation stage of Fairclough's (1989) model and a lack of coherent reasoning for why one analytical method was picked over another. This is a criticism also put forward by Stubbs (1997).

Related to Jeffries' (2010) criticisms listed above, the field of CDA has received further reproach, the most well-cited of which are the criticisms made by Widdowson (1995), who argued that the nomenclature adopted in CDA, specifically the term *discourse*, is so widely used and rarely defined that it is not clear what the term pertains to (Widdowson, 1995). Furthermore, Widdowson (1995) makes the claim that researchers using the term 'discourse' in CDA conflate interpretation and analysis. Additionally, he argues that CDA is not analysis but interpretation and, crucially, that the analyst's interpretation is no more or less valid than anyone else's. He writes:

> There may be reasons for preferring one discourse to another, and if you are ideologically committed you will be inclined to imply that your interpretation of a text is the only one which is valid, that it is somehow in the text indeed, needing only to be discovered, uncovered, revealed by expert exegesis. What is actually revealed is the particular discourse perspective of the interpreter. This may be convincing perhaps, but it has no more authority than any other. To the extent that critical discourse analysis is committed, it cannot provide analysis but only partial interpretation. What analysis would involve would be the demonstration of different interpretations and what language data might be adduced as evidence in each case.
>
> (Widdowson 1995: 169)

Put simply, Widdowson's criticism is that anyone can have an interpretation of a text (e.g. that is contains racist ideology), but it ought to be the role of the expert to provide the means of justifying that interpretation through language data. Moreover, it is easy to see Widdowson's claim about an analyst being "committed" to a particular interpretation as justified when taking van Dijk's (1995: 17) definition of CDA as being 'problem-oriented' into consideration.

In this sense, Widdowson's (1995) and Jeffries' (2010) criticisms of CDA are related as both are about a lack of linguistic description (and therefore objectivity) within CDA, rather than being criticisms of the existence of CDA as a field of study.

Up to this point, I have given considerable space to paying attention to the criticisms of the approach that will make up a significant portion of this thesis, and this may perhaps seem a curious thing to do. However, I have done this for several reasons. The first is that it is important to be aware of these criticisms in order to develop a strategy for addressing them. The second is that discussing these criticisms allows me to define the parameters of what I consider CDA to be. Third, it allows me to define some of the contested terminology used in CDA to avoid any confusion between the terms as they are used in this thesis and how they are used in other analyses of this kind.

To address the first point, I have not relied on any single method of analysis. The analysis reported in this thesis is both qualitative and quantitative, informed by CDA and corpus linguistics. I do not believe that simply using corpus linguistic methods offers a magic bullet for objectivity, but it does provide a set of established linguistic tools that are more objective than simply offering an interpretation of a text. This is by virtue of the fact that much of corpus linguistic analysis is automated by the software[36]. As such, corpus

[36] It is worth stating here that I do not believe any method to be entirely objective. Even in the physical sciences objectivity is just the product of controlled variables and clear parameters. The observation of some 'fact' is still the result of interpretation, e.g. the observation that a

linguistic analysis provides a method for research triangulation. Further to this, I closely controlled the data for qualitative analysis before any textual analysis took place. The data analysed separately from the corpus was sampled using stratified random sampling to avoid any cherry-picking of specific articles or topics (see chapter 4 for more detail on this process). The analytical methods I used to uncover ideology in the data in this thesis were picked because they have been demonstrated to be the best way to reveal the attitudes of the author(s) of texts, as is evidenced by their wide use in CDA research. These methods are used by Fowler et al. (1979) in Critical Linguistics, Fairclough (1989) in CDA and Jeffries' in Critical Stylistics (2010) (albeit under different terms). In response to the potential criticism than there is no set linguistic framework for the analysis of ideology, each analytical method will be applied to each text.

In response to the second point (that is, defining what I consider CDA to be), my own opinion, and what will be reflected in the qualitative part of this thesis, is that CDA falls under the umbrella of Discourse Analysis (*critical* here being a premodifier for an already established area of linguistic enquiry; cf. *interactional* sociolinguistics). This definition is crucial in tying what I am doing in this thesis to linguistics. The object of study in the analysis reported here is language and the methods used to analyse it are taken from linguistics. CDA for me is a linguistic endeavour.

In response to the third point, I need to define what some contested terms in CDA mean in this thesis, namely, 'discourse', 'text' and 'ideology'. I don't endeavour to define discourse and ideology in a way that will be universally agreed upon as I think to do so would be impossible, not least because of the interdisciplinary work being conducted in linguistics now that has, if anything,

---

litmus paper turns blue in alkaline conditions. It is merely the case that the possible interpretations in the litmus paper example are fewer than those on offer to the linguist. See Short & van Peer (1999) for a discussion of objectivity in stylistic analysis.

increased the variety of ways in which the word *discourse* is being used. For the most part, I will neglect to use *discourse* to refer to meaning that exists outside or above the text (as was the sense that Widdowson, 1995 took issue with). Instead I will use 'discourse' to refer to the medium in which a text is presented, for example newspaper discourse, spoken discourse. By *text*, I mean the whole or part of a document that is bound by the concept of cohesion, i.e. coherent meaning and reference (see Halliday & Hassan, 1976). I will use 'ideological meaning', or 'potential ideological implications' to refer to meaning in society caused by the attitudes of the author(s) encoded in the text. I will neglect to refer to 'power' in texts because I see this term as being as problematic as 'discourse'. Both 'power' and 'discourse' imply existing, tangible entities (of vague description) and as such, are top-down interpretations bearing on the data. I subscribe to the belief that all text encodes ideology (as is implied in the Hallidayan metafunctions – particularly the interpersonal metafunction), but it is through the analysis of linguistic structure that ideology is revealed and this ideology need not be inherently negative (as is suggested by 'power', or perhaps more specifically what is often meant by researchers analysing power, *power asymmetry*).

In the next sections, I will describe the two analytical methods taken from CDA that I use in this thesis. These are transitivity analysis (Section 4.2) and naming analysis (section 4.3).

4.2. Transitivity Analysis

It is important that I detail transitivity analysis in the first part of this section as it is transitivity processes that will underpin much of the analytical methods that follow. Moreover, transitivity analysis in its most basic form requires the study of the very basic constituents of the clause (e.g. the noun phrase, the verb

phrase). Transitivity analysis falls under the ideational metafunction and is concerned with the "theory of processes" (Halliday, 2003 [1987]: 127) encoded in the verbal group and how the verbs chosen represent "actions, events and states" (Jeffries, 2010). The types of processes represented have different elements associated with them; for example a material process can take an 'actor' (the doer) and a 'goal' (the done to) whereas a relational process may take a 'carrier' and an 'attribute'. These process types and their related elements are shown in in Table 3.3.2.

Process	**Type**	**Subcategories**	**Elements**
Material	'doing'	Intention (MAI) Supervention (MAS) Event *(if inanimate actor)* (MAE)	Actor, Goal (optional)
Verbalization Process	'saying'	(VP)	Sayer, Verbiage, Goal
Mental	'sensing'	Cognition (MC) Reaction (MR) Perception (MP)	Senser, Phenomenon
Relational	'being'	Intensive (RI) Possessive (RP) Circumstantial (RC)	Carrier, Attribute

TABLE 4.2. TYPES OF PROCESSES IN A TRANSITIVITY ANALYSIS (ADAPTED FROM HALLIDAY, 1973, SIMPSON, 1993 AND JEFFRIES, 2010)

For clarity, I will now explain each process type with examples, starting with material processes.

Material processes are subcategorised into Material Action Intention (MAI), Material Action Supervention (MAS) and Material Action Events

(MAE). The first subcategory is used if the actor is an animate being and the action was done intentionally. The second subcategory, MAS is used if the actor is an animate being but the process was unintentional. The third subcategory, MAE is used if the actor is an inanimate object. See Table 4.3

PROCESS: MATERIAL			
ACTOR	**PROCESS**	**GOAL**	**TYPE**
The student	wrote	her thesis	MAI
The student	lost	her train of thought	MAS
The beer	packs	a punch	MAE

TABLE 4.3. TYPES OF MATERIAL PROCESS

The second transitivity category is Verbalization Process (VP). This process type includes any action that describes communication. The elements the category takes include the sayer (the actor 'doing' the communication), the verb describing how the message was communicated (e.g. said, shouted), the verbiage, or what was said, and the target, or who the verbiage was addressed to. As is shown in Table 4.4 below, the verbiage and the target are not essential elements in the verbalization process.

PROCESS: VERBALIZATION			
SAYER	**PROCESS**	**VERBIAGE**	**TARGET**
The tutor	said	the assessment was due	Ø
The tutor	shouted	Ø	Ø
The tutor	said	yes	to the students

TABLE 4.4. TYPES OF VERBALIZATION PROCESS

The next category is mental processes. This category is concerned with how an actor perceives, processes or reacts to something in the world. This category contains three subcategories. The first is Mental Cognition (MC), which

includes verbs of cognition such as *think*, *realise*, etc. The second subcategory is Mental Reaction (MR), which includes verbs that relate to emotional states such as *loving* or *hating*. The third subcategory is Mental Perception (MP), which relates to verbs that describe sensing something in the world, e.g. *hearing* or *seeing*. See Table 4.5 below.

PROCESS: MENTAL			
SENSER	**PROCESS**	**PHENOMENON**	**TYPE**
The public	realised	the referendum result was void	MC
The student	despaired	at the result	MR
The Prime Minister	saw	the protests	MP

TABLE 4.5. TYPES OF MENTAL PROCESS

The final category is relational processes[37], which relate to processes describing being or having, e.g. using the copular verb, or describing some attribute of the entity being discussed. Relational processes have two modes that either express a process of identifying or attributing (Halliday & Matthiessen, 2004: 244). Whether a clause is identifying or attributive can be expressed in three types of relational process. The first is the Relational Intensive (RI) subcategory, which expresses an *x is y* relationship; the second is the Relational Possessive (RP) subcategory, which expresses an *x has y* relationship; and the third is the Relational Circumstantial (RC) category, which expresses an *x is at/in/on y* relationship (Halliday & Matthiessen, 2004: 440; Simpson, 1993: 91-92). Within these categories, SFL researchers use further subcategories for elements that refer to the types of information conveyed by relational processes at the level

[37] This short section is necessarily long in comparison to the other two process types. This is because Relational processes are much more complex in their subdivision, and rely much more on the theory underpinning SFL. Moreover, I have yet to find a text that coherently sums up relational processes without first stating that the process is "complicated". Given that something being complicated does not seem to me to be a satisfactory reason for not detailing what it is, I have tried to explain what Relational processes are here, drawing where necessary on SFL theory.

of the lexical item. For example, in relational clauses that identify someone or something, the participants may be described as the *identifier* and the *identified*, and when something or someone is being assigned a status of some sort, the participants may be described as the *possessor* and the *possessed*. The complexity of the Relational Process has meant that some researchers (particularly those working outside of SFL and who use transitivity as a method rather than a theory), have used simplified versions of the Relational category, simply referring to all participants as *carrier* and *attribute* (Jeffries, 2010; Simpson, 1993), or only adopting the *carrier/attribute* and *token/value* elements without describing the nature of the process, i.e. possessive, identifying (Canning, 2013). As can be seen by the table below, for theorists of SFL, the attributive mode is concerned with describing the attributes of an entity that classify it as part of a group, whereas the identifying mode is about relating two entities in some way. A helpful way to distinguish between the two modes is to note whether the definite article (identifying) or indefinite article (attributive) is used. See Table 4.6 for a summary.

Intensive	Emily is a poet	attributive
	Emily is the poet	identifying
Possessive	Emily has a piano	attributive
	The piano is Emily's	identifying
Circumstantial	The meeting is on Friday	attributive
	The time of the meeting is Friday	identifying

TABLE 4.6. TYPES OF RELATIONAL PROCESS (TAKEN FROM HALLIDAY (2004: 239))

Table 4.7 below shows examples of the three process types with some expanded information about the participants in the process.

PROCESS: RELATIONAL			
ATTRIBUTIVE PROCESSES			
Carrier	*Process*	*Attribute*	
Dan	is	a brilliant supervisor	RI
Possessor	*Possession*	*Possessed*	
Jack	has	a green bike	RP
Carrier	*Process*	*Attribute*	
The grammar book	is	in the bin	RC
IDENTIFYING PROCESSES			
Token: Identified	*Process*	*Value: Identifier*	
Lynn	is	my mother	RI
Token: possession	*Process*	*Value: possessor*	
The thesis	was	the student's	RP
Token: identified	*Process*	*Value: identifier*	
Sunday	is	tomorrow	RC

TABLE 4.7. ATTRIBUTIVE AND IDENTIFYING PROCESSES WITHIN THE RELATIONAL PROCESSES CATEGORY

Although Jeffries and Simpson do not state this explicitly, the reason the model they adopt (Table 3.3.2) is simplified is that they observe *delicacy* in a lesser degree than I do here. Halliday defines delicacy as a "scale of differentiation, or depth in detail" (Halliday, 2002: 58) with the primary degree referring to "categories of structure and class" (Halliday, 2002: 58) and the secondary degree (although Halliday says there is no limit to delicacy [Halliday, 2002: 405]) referring to the categorisation of lexical choice.[38] It seems to be the case that delicacy refers in the primary form to the necessary constituents of a process. In its secondary form, delicacy refers to the categorization of lexical choices or the nature of lexical networks, e.g. in a RC process, the lexis contained will usually pertain to place, time or manner. We can conceptualise delicacy at this level as being syntagmatic (Hasan, 1987) and as related to the

[38] Put simply, *delicacy* relates to the level of abstraction (Williams et al, 2017).

Firthian concept of collocation, whereby lexical choice is governed by possible semantic structures that exist earlier in clause (what Martin, 2016: 45) terms "expectancy relations". The belief that semantic meaning is structured and governed by rules just like those on the syntagmatic level led Halliday to state that "lexis is the most delicate grammar" (1961: 267).

Relational processes are used in varying depths of "delicacy" (Matthiessen & Halliday, 1997: 29) or "the degree of detail in which a structure is specified" (Butler 1985: 19) by researchers in SFL (specifically for my purposes here, the description of more realization options within RI, RP and RC processes). For instance, Bloor & Bloor (1995: 127) and Butler (2003: 429) discuss the elements "identifier and identified" in reference to Relational processes, Halliday and Matthiessen (2004: 173) refer to token/value elements, and Fawcett (1987: 161) further develops the 'carrier' element. For clarity, the possible elements or participants a process type can take are shown in 3.3.2.

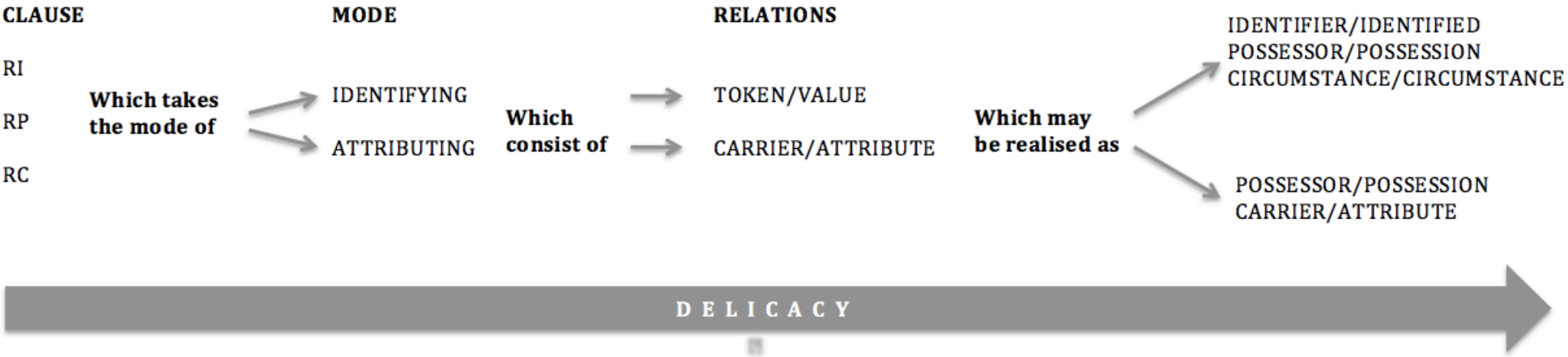


FIGURE 4.2. A BREAKDOWN OF RELATIONAL PROCESSES AND THEIR PARTICIPANTS

Generally speaking, my aim in this thesis is not to account for the description of all variations of a clause (or even all the possible lexical choices the producer of a text could have made), but rather the general patterns contained in the clauses for the purposes of analysing potential ideological effects. As such, my use of the framework is one of application (i.e. it is used here as a method) rather than for the purpose of furthering the grammar itself. With this in mind, I have adopted an intermediate approach to the analysis of relational processes, something between the simplified model adopted by Jeffries (2010) and Simpson (1993), and the full model used in formal SFL in which the possible subcategories of carrier and attribute appear to vary considerably by delicacy. My reason for taking this course of action is that I believe that simplifying the model to the degree that Jeffries (2010) and Simpson (1993) do would potentially limit some interesting findings within the data – for example, whether mental illness in relational clauses is expressed as an attribute, an identifier, a possession or a value of a person.[39]

As can be seen above in Table 3.3.2 and the examples for each process type, the value of transitivity analysis is that it is combines a syntactic and semantic description of language, which allows for the analysis of ideological effects. Moreover, it is concerned not just with the properties of the verbal or nominal group, but also with the meaning potential of the sum of the two. With this in mind, it is important to note that in any model that combines the description of a heavily structured system (syntax) and a system that necessitates greater interpretation of co(n)textual meaning (semantics), categorisation may not be straightforward. Matthiessen & Halliday (1997) recognise this and use the lexical item 'make' to demonstrate how this item can be coded as a material process (as a synonym of 'produce') as well as a relational process (to mean 'cause to be') (Matthiessen & Halliday (1997: 16). Such intermediate cases within the relational category have been identified in detail by Cerban (2008) and,

[39] With this said, it seems to me that the inclusion of a greater number of subcategories provides more variables than is useful in some cases. I refer here to the principle of Occam's Razor, which states that "entities should not be multiplied unnecessarily." (Heuristic of unknown origin).

generally speaking, it appears to be the case that the greatest number of intermediate cases occur between the Material and Relational categories only. The complex structure of SFL, which leads to fuzzy cases like those identified by Cerban (2008), mark a potential weakness of SFL – that its complexity eschews clear application in some cases (see the principle of Occam's Razor in footnote 39). SFL is useful because, compared to other, more parsimonious models of grammar, e.g. Chomskyan minimalism, it takes into account (or perhaps more accurately recognises that it takes into account) semantic meaning, which necessitates a more complex theoretical architecture (Williams et al., 2017).

In line with Jeffries and McIntyre, I argue that "[t]he key to dealing with the problems of categorisation is not to treat transitivity types as categories at all, but as points of reference on a continuous plane of meaning" (Jeffries & McIntyre, 2010: 74). Wherever relevant, in problematic cases, I will provide a commentary detailing why certain clauses were categorised in the way that they were. Furthermore, my decision to adopt an intermediate model for relational processes provides a set of diagnostics to aid the categorisation process. For example, the participants in a relational attributive process can be reversed (with changes to the grammatical role of the subject) but those in an identifying process generally cannot.

IDENTIFYING

IDENTIFIED	VERB	IDENTIFIER
Lynn	*is*	*my mother*
IDENTIFIED	VERB	IDENTIFIER
My mother	*is*	*Lynn*

ATTRIBUTIVE

POSSESSOR	VERB	POSSESSED
Jack	*has*	*brown hair*
* *Brown hair*	*has*	*Jack*[40]

[40] I feel obliged as a Yorkshire [wo]man to point out here that whilst being ungrammatical, this type of usage is acceptable in some British English varieties, such as the Yorkshire dialect, e.g. '*lovely eyes*

In this sense, the model I have adopted can account for the grammatical and semantic description of language albeit with simplified delicacy.

4.3. Naming Analysis

Naming analysis refers to the ways in which participants or entities are named or "packaged-up" (Jeffries, 2010: 19). Naming is "a powerful ideological tool [and] […] an accurate pointer to the ideology of the namer [because] different names for an object represent different ways of perceiving it" (Clark, 1992: 209). The analysis of naming practices is concerned with the nominal group and how the basic unit of the nominal group, e.g. a noun, is chosen over another noun, or how information about the noun is modified, e.g. through adjectival and prepositional phrases or complements that convey extra information about the thing being discussed.[41] Sentences (1) and (2) illustrate these two types of naming practice via fictional newspaper headlines:

(1) *A woman* was attacked while out walking
(2) *A blonde in a miniskirt* was attacked while walking

In (1) we see that the entity contained in the nominal group has been represented using the lexical item 'woman'. This is a fairly unsurprising lexical choice given that the purpose of crime reporting is to give some information about the crime, the victim and if possible, the perpetrator. The choice to use 'woman' in this headline over, for example, 'a person' may be guided by the nature of the attack (for example, if the motivation for the attack led police to believe that women were at greater risk than

has my mum'. (This usage is one of the lesser known linguistic links between Doncaster and the Dagobah system).

[41] Verbal processes may also be realized as nouns as is the case in nominalized forms or deverbal nouns.

men). Furthermore, other naming options may not be possible for legal reasons or for reasons of confidentiality, for example, using the person's full name

In (2) the nature of the naming practice is different from that in (1), as the noun has changed ('woman' to 'blonde' – where blonde is a meronym[42] denoting a blonde woman) and is also post-modified by a prepositional phrase ("in a miniskirt"). This sentence is much more clearly ideological than that shown in (1). The reasons for this are multiple. For example, the choice to use 'blonde' (which has connotations of attractiveness and sexuality – particularly a conscious effort to be conventionally attractive [cf. dyed/platinum blonde])[43] over 'woman' foregrounds the appearance of the woman and provides unnecessary information. It is hard to see any purpose for including this information about the woman other than to suggest that her appearance was in some way linked to the attack. This is furthered by the choice to include what the woman was wearing (an item of clothing that displays her legs) in a post-modifying prepositional phrase. As a result of these linguistic choices, what looks on the surface to be a simple description of the woman and her clothing allows for the presentation of covert, ideological meaning – specifically, that the woman was not innocent in the attack or that by simply wearing a particular item of clothing and having a particular hair colour, the woman is not blameless for the attack. Naming practices then, have an important ideological function in that they reveal information about an author's ideological viewpoint. Moreover, naming practices can reveal the connotations that particular lexical items have in society. By way of qualifying this final point, consider sentence (3).

(3) *A grey in jeans* was attacked whilst out walking

[42] In lexical semantics, meronymy is a sense relation whereby part of something is used to describe the whole of something, e.g. "boots on the ground" to refer to soldiers.
[43] The top 100 collocates of 'blonde' in the BNC include 'dumb', 'attractive', 'dyed', 'buxom', 'sexy', 'legs', 'chest'.

Sentence (3) retains the structure of the original sentence but the nominal group has been altered to include lexical items that do not carry the same associations as those in (2) by way semantic substitution test. Most people (at least in an Anglo-western context) would struggle to parse (3) because *grey* as a referent is not conventionalised[44]. Moreover, the relevance of what the woman was wearing is more easily questioned because 'jeans' do not carry the same connotations as 'miniskirt' despite the fact that what clothing the woman was wearing is in both instances irrelevant.

The sentences I have presented to illustrate naming practices so far may seem fairly extreme. However, existing research into naming practices has analysed headlines taken from national newspapers that use very similar naming practices to the ones used here. For example, Clark (1992) analysed newspaper articles that appeared in *The Sun* between 1986-1987 that reported on male to female violence. Clark found that female victims were referred to as (amongst many other referents) 'blonde', 'prostitute', 'divorcee', 'unmarried mum' and 'Lolita'. Clark found that the terms used to refer to the victims of violence served to label them without individualising them, e.g. "Sex-starved squaddie strangled blonde" (Clark, 1992: 218). Such constructions, Clark argues, portray the victim "in terms of her sexual attractiveness as something which any man, especially one named as 'sex-starved', could not help responding to" (Clark, 1992: 218). Moreover, such constructions leave little room for any sympathy for the victim; rather the naming practices hold the women responsible.

[44] I believe it to be the case that this example is not just deviant for the reasons I have pointed out above, but also because of the maxim in Anglo-western culture that stipulates that we have to respect our elders, meaning that, as a result, referring to an elder by their physical characteristics, e.g. a 'grey', would be deemed rude. Conversely, referring to (young) women by their physical characteristics seems to be an accepted feature of mainstream society. The closest examples of hair being used to identify a man that come to mind are the set phrases: 'talk, dark and handsome' which seems not to have the negative connotations that 'blonde' does, and 'silver fox' which is also positive and is usually used to refer to middle-aged men only. The only potentially negative term is 'baldy' which only occurs 12 times in the entire BNC with only 4 instances referring to men (the referents in the other instances were inanimate objects, animals, unknown, or women).

So far, the examples I have explored have relied on information about the entity being foregrounded (e.g. physical features) which works to background other features that may not fit with the representation the author of a text wishes to convey. In foregrounding certain aspects of a person or entity, creators of texts are exploiting the unavoidable nature of nouns (i.e. that they name entities) in order to establish a person or thing in the world according to their opinion. As a result, it is hard to unpick the ideology a nominal group contains because the very nature of naming something presupposes its existence (especially in less extreme cases than the ones I have offered so far). Jeffries writes of this phenomenon "the nominal component […] does not form the *proposition* of the clause or sentence but instead labels something that is thus assumed (technically, *presupposed*) to exist" (Jeffries, 2010: 21, original emphasis). Consider for example, sentences (4) and (5) where the ideological content encoded is less obvious.

(4) When I went to the hospital I saw a lovely *lady doctor*

(5) Mr and *Mrs May* walked out of Number 10

The nominal groups of interest in (4) and (5) are much less obviously ideological than the previous examples and for some people (particular, perhaps, older people) may not appear to be marked at all. This is due to the fact that both naming practices shown in (4) and (5) reflect ways of referring that are entrenched in British society, and which draw upon traditional views about the roles of women in the workplace and in marriage. For example, in sentence (4) the choice to pre-modify 'doctor' with 'lady' is indicative of the view that most doctors are men. This is despite the fact that there is no formal semantic reason why 'doctor' can't be used to refer to a woman. Furthermore, the noun phrase 'doctor' does not require pre-modification to be syntactically complete. So, if the semantics do not encode maleness and the syntax does not require pre-modification, then the need to include 'lady' has to be an

ideological decision[45]. As such, its usage provides insight into what society (or more specifically people in a society) deems the norm to be. In sentence (5), the ideology encoded in the naming of the referent is harder still to identify. The reason for this is that it is more removed from language than that shown in (4), i.e. there is no syntactic reason for this naming practice to be marked. The reason (5) is of interest is that the referent named as 'Mrs May' (who might also be named as 'the former Prime Minister of the United Kingdom or 'destroyer of hopes, dreams and fundamental human rights' according to your ideological viewpoint) is named in relation to another person, specifically her husband. The choice to name Theresa May in this way is indicative of how cultural practices affect naming options, e.g. it is tradition (and still very much conventional) in an Anglo-Western context for a woman to take her husband's name after marriage[46]. The decision to name Theresa May in relation to her status as a wife and not by her status as the former Prime Minister demonstrates again how naming can be used to foreground aspects of a person or entity (her status as a wife) and background others (her status as a head of state)[47]. The decision to use 'Mrs May' and not 'Theresa May' or 'the former Prime Minister' could be a useful strategy for conveying traditional values about marriage and the family unit, which may be in line with Conservative Party ideals. Conversely it may be a sexist strategy to downplay May's status because of her gender (it may also, of course, be entirely unconscious and unintended to have an ideological effect; but the usage still generates an ideological meaning). Moreover, the structure of (5) does allow for nominal apposition, where two grammatical elements (usually nouns) placed together identify the same referent using different descriptions (6). And, the same proposition could be expressed without placing 'Mr May' in subject position, see (6) and (7).

[45] This is also good evidence to suggest that semantics and syntax alone are not enough to describe language; at the very least we have to recognise a semantics/pragmatics interface.
[46] It will be interesting to see whether same-sex marriages start to change this convention.
[47] As an interesting aside, a search for "Mr and Mrs May" in the News tab of Google yields around twice as many hits than "Mr and Mrs Cameron".

(6) *Mrs May, the Prime Minister*, and Mr May walked into Number 10

(7) *The Prime Minister, Theresa May* walked into Number 10 with her husband

It can be said, then, that in a general sense, the analysis of naming practices offer two vantage points from which to analyse current and societal views because, (i) naming shows how people are limited in their naming choices because of societal norms (e.g. 'Mr and Mrs') , and (ii) society is constructed based on how people name things within it (e.g. the view that the most important elements of a women are her physical attributes, such as in the 'blonde' example). Point (i) here refers to the deterministic view that changing the labels in a language will change the way that people think about that thing. Point (ii) refers to a social constructivist perspective (See Chapter 2) on the relationship between language and society, where a change of the status of something in society results in a change in language. However these two frameworks are not distinct, as Mills (2003) writes: "language items both affirm and contest the status quo and changes in social structures necessitate the development of new vocabulary" (Mills, 2003: 88-89). Page (2003) shares this sentiment, writing of naming practices in media discourse:

> This relationship with social reality is complex and dialectical. At one level, as a cultural artefact, media discourse is part of social reality itself. However, the relationship is not static, but the discourse operates within particular social contexts, and is said to both be affected by and able to affect the power relations embedded therein
>
> (Page 2003: 559-60)

Both the deterministic view and the social constructivist view have relevance to the analysis of naming practices as both imply that naming practices (as part of language) can offer insight into the identity construction of the human participants named within the text from an emic perspective (i.e. how a person constructs their own

identity through language) and an etic perspective (i.e. how a person's identity is contrasted by others through language). These insights are invaluable in this thesis because many of the terms used to label mental illness and people with mental illness are contested, still emerging and, most importantly, have real world consequences for the individual's sense of self and access to resources. Moreover, the diachronic nature of the corpus used in this thesis allows for the analysis of naming practices across time, for instance the use of 'person-first' language (e.g. 'a person with schizophrenia' instead of 'a schizophrenic') or the avoidance of certain descriptions for people with a mental illness (e.g. 'sufferer' to 'patient' or 'service user'). I conduct a naming analysis of these aspects of the corpus in Chapter 7.

4.4. Conclusion

In this chapter I outlined the analytical methods I use in this thesis that are taken from CDA. In section 4.1, I described CDA from its beginnings in Critical Linguistics to contemporary research like that conducted in this thesis, that combines methods from CDA and corpus linguistics. In Section4.1, I also reviewed the existing research into ideology in language making the argument that some of the existing research in CDA is open to criticisms of overreliance on interpretation. I also discussed the notion that combining methods from CDA with corpus analysis provides a magic bullet for objectivity in research exploring ideology in language. Along with Chapter 3, then, this Chapter outlines the analytical methods I use in the analysis I report in this thesis. Specifically, I conduct naming analysis in Chapter 7 to explore the salient ways that people with mental illness are named in the press. Furthermore, I conduct a transitivity analysis to explore how the press describe people as 'having' or 'experiencing' mental illness in Chapter 8.

In the next Chapter, I describe the process of constructing the MI 1984-2014 corpus.

5. Corpus construction

This chapter outlines the method for data collection used in this thesis, focussing on the process of compiling the Mental Illness 1984-2014 corpus and its related subcorpora. In the following sections, the various stages of corpus construction will be detailed. In Section 5.1, I focus specifically on what role search terms play in the process of building corpora, but also the role they play in the analysis stages. In Sections 5.2 - 5.3, I describe how the search terms for the Mental Illness 1984-2014 Corpus were generated and describe the sampling frame for the data collection. Section 5.4 discusses the corpus cleaning procedure. In section 5.5, the procedure for constructing subcorpora is outlined. I discuss the practical issues arising from the subcorpus construction procedure in Section 5.6. Section 5.7 shows how the final corpus is representative of the target population using some basic frequency and dispersal tests. Section 5.8 discusses and responds to any methodological caveats before I conclude this chapter in Section 5.9.

5.1. Search terms: interpretatively neutral?

Critical corpus linguistic studies rarely discuss the role of semantic content in the compilation of search terms, preferring instead to focus on linguistic form[48]. Discussion of search terms tends to focus on (i) general issues of how analysts can use lemmatisers and basic regular expressions (e.g. wildcards[49]) to ensure that all possible derivations of a root word are captured effectively (e.g. 'immigr*' to capture 'immigrant' and 'immigration'); and (ii) whether the search parameters effectively capture the medium the researcher wants to represent, e.g. tabloids or broadsheets

[48] My argument here is mainly in reference to studies that make use of data from press publications.
[49] A wildcard (e.g. *,!) is a regular expression used in searches to allow the search to return particular strings of characters.

(see Gabrielatos & Baker, 2008). Discussion about search term syntax is arguably due to the fact that many of the analytical methods used in corpus linguistics are continually being developed and, as a result, shared techniques to capture all instances of a lemma when building corpora are still useful and relevant in the research community. This aside, the focus of the process of compiling search terms in studies that adopt CL and CDA methods tends to be on discussion of the data sources used, such as tabloids and broadsheets, and their historic position on societal issues such as Islamaphobic views (Baker et al. 2013), rather than how to best capture the language used to represent that societal issue, or the medium the analyst wants to represent. Indeed in their 2013 article, Baker et al. write that their search term was developed "using trial and error" (2013: 259). This approach to data collection suggests that instead of letting the data reveal the textual practices at play within it, analysts are collecting data with *a priori* research objectives[50] and aiming to validate or invalidate particular hypotheses (e.g. 'tabloids are more negative in the representation of immigration than broadsheets' or 'right-wing newspapers stigmatise Islam'). This practice has been described in the CDA literature as 'problem-oriented social research' (Baker et al, 2008: 279), which raises an issue for corpus linguistic informed CDA studies generally: how does a researcher represent a population for the purposes of studying objectively the ideologies in that population? It has to be the case that search terms are used to build a corpus that contains instances of the type of language that the analyst wishes to study, to allow for cross analysis of related terms (see, for example, Baker et al.'s, 2008 study on 'immigrant', 'migrant', 'asylum seeker', etc.). However, researchers doing this could be accused of lacking the objectivity that the

[50] Arguably, this is particularly true of fields like CDA where identifying ideologies is the aim of analysis rather than a potential research finding. This is the result of a difference in focus between traditional corpus linguistic ideas (i.e. representing a type and /or domain of language such as spoken, written, scientific or fiction) and the ideas of CDA (i.e. looking for particular ideologies within a variety). Put simply, traditional CL analysis was about mapping patterns of language without any focus on ascribing value to usage. CDA is inherently about ascribing value to usage. This difference in focus is a result of the tendency in CDA to observe the notion of 'discourse' over the notion of linguistic 'tokens'.

field of corpus linguistics was developed to afford them. The issue of objectivity in CDA has been criticised from another perspective also; Koller and Mautner (2004) have argued that the texts used in CDA (and relatedly, critical corpus linguistics) have been cherry-picked (consciously or not) in order to give positive evidence for a discourse or ideology, rather than letting discourses be identified in a bottom-up approach:

> The hidden danger is that the reason why the texts concerned are singled out for analysis in the first place is that they are not typical, but in fact quite unusual instances which have aroused the analyst's attention.
>
> (Koller and Mautner, 2004: 218)

Furthermore, Sinclair writes of researchers constructing corpora with specific insight into its content that there is "a danger of a vicious circle arising if they construct a corpus to reflect what they already know or can guess about its linguistic detail" (Sinclair, 2005: 1). In fact, for Sinclair, a corpus should be built without any insight into what the texts contain (Sinclair, 2005). However, building a corpus without knowing what the corpus contains poses problems for the critical corpus linguist working outside the theoretical tradition of corpus linguistics, and whose research necessitates the construction of a corpus based on its internal criteria; that is, the topic the text discusses. Moreover, the very construction of specialised corpora requires specific words to be present in order for potential constituent texts to be deemed relevant for inclusion. What is deemed good practice in corpus construction when using corpus linguistics as a method in CDA research, then, is clearly different from that advocated by Sinclair (2005). This is not to say that the corpora created with these research aims are any less valid than those created observing the criteria Sinclair (2005) sets out. However, what it does mean is that researchers building specialised corpora ought to pay attention to the role of search terms in corpus construction as the search terms play a huge role in the analysis – because they constitute the data. In Section 2, I draw

attention to this otherwise neglected topic in the discussion of (critical) corpus linguistic methods, which is the inherently interpretative process of collecting data through search terms. I suggest that paying greater attention to the compilation of search terms could go some way towards addressing criticisms about objectivity in critical discourse analysis from the outset, simply by making the analyst aware that words are in themselves social constructions; i.e. to search for a word in a corpus, the analyst is recognising (tacitly or not) that the word means something in society[51]. This brings me to my next point. In many CDA projects, researchers do not acknowledge that the search terms used are in themselves a form of interpretation (or at least work to prime an interpretation of the data before it is analysed). If what we are trying to do when building corpora is model a population, then identifying key search terms as a way to model that population necessarily indicates that a degree of interpretation is taking place before the formal stage of analysis begins. Instead of recognising this, however, some researchers have even suggested that corpora built with specific research purposes in mind facilitate objective analysis because corpus linguistic methods 'enable the researcher to approach the texts (or text surface) (relatively) free from any preconceived or existing notions regarding their linguistic or semantic/pragmatic content' (Baker et al, 2008: 277). Clearly this is a problematic line of argument when corpora are built with the very intention that they contain specific semantic and pragmatic content. The techniques used to analyse the semantic and pragmatic content of the corpus may be (relatively) neutral; the content of the corpus being analysed is not.

To use an analogous example from another sub-field of linguistics, I propose that critical corpus linguists should view corpora constructed to represent a particular societal issue in the same way that Conversation Analysts view transcripts of conversation. A transcript is not the data. Rather, it is an interpretation of the data and necessarily contains some interpretative decisions, e.g. the decision to transcribe

[51] This is to say, the word means something in society more than simply being a noun or a verb, or as being a feature of a particular discourse type, e.g. spoken discourse.

laughter with a particular vowel quality. Just like transcripts, so too do search terms encode interpretation. For this reason, researchers constructing specialised corpora using search terms (and particularly those that target a potentially contentious social issue) ought to be able to ask themselves questions such as: Why are the words chosen significant? Why did I pick these words from the other words available? Has my own position on this societal issue led to terms being missed or added without good reason?

To view the search terms as almost an aside to the corpus construction process should be a concern when one considers that the whole analysis is predicated on the data elicited using those search terms.

5.2. Search terms and sampling frame for the MI 1984-2014 corpus

In this section I outline the process of determining search terms for the MI corpus in a way that avoids the pitfalls I identified in the previous section. I begin by explaining the rationale for the search terms.

5.2.1. Generating search terms

In order to accurately represent the population I was sampling from, I first turned to the UK based mental health charity *Mind*, given that "An obvious starting point for the compilation of a query is lexis denoting the entities, concepts, states, relations or processes that are to be investigated" (Gabrielatos, 2007: 6). The reason for choosing the *Mind* website over other sites offering information on mental health and illness was that it features an A-Z of mental health (*Mind*, 2018). This provided a near exhaustive list of mental health terms that included accessible terms for illnesses. For example, the terms featured were referenced by their common names, not by medical jargon. The preference for using common names for illnesses in the press over medical

jargon was something I observed in a small-scale pilot study I conducted during the search term generation process. For this reason, I made the informed decision to use *Mind*'s A-Z over the terms listed in the *Diagnostic Statistical Manual* (Fifth Edition) (DSM-V) which offers an exhaustive list of mental illnesses and symptoms aimed at practitioners. A further reason for opting to use *Mind*'s A-Z over the DSM-V was that the typology of mental illness included in the DSM-V has to be very broad and include specific information because it is used by practitioners. As a result, the illnesses it includes are more for the purposes of diagnosis than description. Furthermore, given that my interest in the current study was to track the social construction of mental illness by people who may not be medical experts in the field of mental illness, I felt using accessible terms for the illnesses would better reflect the view of mental illness in society (i.e. the view of the non-expert). Moreover, I made the decision not to include pejorative or euphemistic terms for referring to mental illnesses in my search terms (e.g. nutcase or bonkers) in order to focus my analysis on medicalised terms. There is, of course, the high probability that such terms are included in the final corpus despite this methodological decision.

Section 5.2.2 provides some information about *Mind* and the A-Z of Mental Health.

5.2.2. Mind

Mind is a mental health charity founded in 1946. The charity covers England and Wales and works with local authorities to help people suffering with mental ill health and their families to access services. On their website, *Mind* state:

> We provide advice and support to empower anyone experiencing a mental health problem. We campaign to improve services, raise awareness and promote understanding. **We won't give up** until everyone experiencing a mental health problem gets **support and respect**."

(Mind, 2019b; original emphasis)

In addition to working with individual service users, *Mind* is also associated with several public-facing schemes to raise awareness of mental health in society. They have *Mind* celebrity spokespeople (Stephen Fry is the President of the charity) and work across platforms to reach different target groups. A recent example of this was *Mind*'s digital ambassador scheme, for which they made the YouTuber Zoe 'Zoella' Sugg [52] the charity's first digital ambassador, to launch their #DontPanicButton campaign aimed at raising awareness of anxiety and panic attacks in young people.
In addition to their work raising awareness of mental health through celebrities, *Mind* also have a 'Media Office' section of their website that allows users to access spokespeople to talk about mental health. As part of their media work, *Mind* also offer a service on how to report on mental health in which they give a list of the Press Complaints Commission Code of Practices on reporting on mental health and how to report on specific issues such as suicide, violence and eating disorders. This is also a service offered by *Rethink Mental Illness*, another charity that *Mind* collaborate with.

Knowing this, it is clear that *Mind* have expertise on the issue of mental health but also how societal perceptions of mental health are created, maintained or changed by language. For this reason, I used *Mind* as the starting point when compiling search terms for the Mental Illness 1984-2014 corpus.

5.2.3. Using the Mind 'A-Z of Mental Health' to build search terms

The *Mind* 'A-Z of mental health' lists all the terms *Mind* see as the key terminology of mental health, including diagnosable conditions and symptoms. The criterion for

[52] A YouTuber is a person who creates and features in YouTube videos. Zoella's YouTube channel has nearly 12,000,000 subscribers. A video posted on her channel titled 'Dealing with Panic Attacks & Anxiety' has been viewed over 4,000,000 times.

inclusion in my search terms was that the term had to relate to a diagnosis but not be a symptom of a diagnosis (e.g. panic attacks are a symptom of the diagnosable condition anxiety). The rationale for this was that articles on the topic of anxiety would contain 'panic attack' and therefore the term would be captured by the search terms anyway. Moreover, many of the symptoms related to multiple illnesses, which meant that should an article on anxiety not capture 'panic attack', another search term may. Another reason for not including symptoms was that some symptoms could introduce noise into the corpus. 'Stress' is an example, as it does not necessarily refer to mental illness.

Restricting the terms in this way raised an important issue about how mental health disorders are grouped, as although the difference between symptoms and diagnosis looks to be clear, the waters are muddied by conditions like dissociative disorder, which has several sub-diagnoses within it but whose symptoms (dissociation) are also symptoms of other disorders such as schizophrenia, bipolar and borderline personality disorder. The complex phobias (social phobia and agoraphobia) were included, but specific phobias were not (*specific* and *complex* are terms taken from *Mind*).

Once the terms for specific mental illnesses were collected using the *Mind* A-Z, general terms to refer to mental illness and health were added to ensure that articles reporting on the issue generally without reference to specific illnesses would be captured. These include 'mental illness', 'mental health', 'mental ill health', 'mentally ill', 'mentally un![53]'. In addition to these general terms, 'autism' was also added as a search term. The rationale for including autism was that people with autism are more likely to experience mental illness than members of the general population (National Autistic Society, 2019), and yet despite this, "the mental health of autistic people is often overlooked" (National Autistic Society, 2019). For this reason, including autism allowed for the capture of articles that may have not been captured otherwise, and

[53] Recall that '!' used in search terms is a wildcard.

also provided the opportunity to explore the overlooked relationship between autism and mental illness. It is important to state at this point that there are very many other groups of people in which the prevalence of mental illness is higher than average (e.g. those below the poverty line or those living alone). The reason for including 'autism' and not these other groups however was due to the fact that 'autism' has a clear lexical item associated with it while these other groups do not. Including search terms that are not immediately relevant to the other search terms but are related can offer insight into the topic being studied because they can identify articles that discuss mental illness without using terminology directly related to it. Gabrielatos (2007) writes on this topic in reference to a study into the words 'refugees' and 'asylum seekers':

> if an article reports on or discusses issues related, directly or indirectly, to refugees or asylum seekers, these two groups may not necessarily be referred to explicitly. If, however, the query string includes as many other terms as possible referring to the same or similar groups, then it is expected to capture a large proportion of those articles in which the groups in question are not mentioned explicitly
>
> (Gabrielatos 2007: 8)

This final search term collection process resulted in 26 lexical items/phrases. The terms selected from *Mind* also cover a broad range of disorder types as described in the DSM-V. Figure 5.1 shows the search terms grouped by disorder type:

General Terms
mental illness, mental health, mental ill health, mentally ill, mentally un!

Trauma Disorders
Post-traumatic Stress Disorder

Autism Spectrum Disorders
Autism

Dissociative Disorders
Dissociative identity disorder

Bipolar Disorders
Bipolar
Mania
Hypomania
Hypermania

Obsessive Compulsive Disorders
Obsessive Compulsive Disorders
Body Dismorphia

Personality Disorders
Personality Disorder
Borderline Personality Disorder

Psychotic Disorders
Psychosis
Schizophrenia

Anxiety Disorders
Agoraphobia
Social Phobia

Eating Disorders
Bulimia
Anorexia
Binge eating disorder

Depressive Disorders
Depression
Seasonal Affective Disorder - (depression with seasonal pattern)
Postnatal Depression

FIGURE 5.1. LEXICAL ITEMS GROUPED BY DISORDER TYPE AS CATEGORISED IN DSM-V.

Once these terms were decided upon, the lexical items were prepared for incorporating into a search term by using wildcards to ensure that all variants of the lexeme would be captured in a search. This included starting with the nominalised form of the word, e.g. 'psychosis', and working through the different parts of speech it could take, e.g. adjectival forms such a 'psychotic'. Some of the lexical items have a nominal form that can perform an adjectival function, e.g. 'he was bipolar', 'he has bipolar'. For others, the adjectival form was needed, e.g. 'he was psychotic' (c.f. 'he has psychosis'). Issues such as these were alleviated by using wildcards in the specific

terms e.g. 'agoraphobi*'. However, wildcards were not used in lexemes such as 'anxiety', since high frequency words like 'anxious' would create noise in the corpus, as the term is used most commonly outside of a mental health context. . Similarly, with terms like 'seasonal affective disorder' that have common initialised forms, i.e. 'SAD', only the full term was searched for to avoid further noise. Any initialisms in the search terms only refer to the mental health disorder, e.g. 'BPD' for borderline personality disorder.

5.2.4. Final search terms

As far as possible, the final terms comprising the search term were designed to capture a variety of spelling variants of words, (e.g. 'posttraumatic stress disorder', 'post-traumatic stress disorder'), acronyms of conditions, (e.g. 'OCD'), and a variety of possible ways to refer to a condition (e.g. 'bulimia', 'bulimia nervosa', 'eating disorder'). The design of the overall search term means that it captures a minimum of 49 ways of referring to mental health phenomena. These include diagnosable illnesses (e.g. hypomania), assessments of states of mind (e.g. mentally unstable) and mental health literature (e.g. mental health act). Table 5.1 shows the breakdown of the constituent search terms and the final overarching search term.

Condition or lexical item	With wildcard (if added)	What it will capture
mental health	(mental health!)	mental health, mental health act
mental ill health	(mental ill health)	mental ill health
mental illness	(mental illness!)	mental illness, mental illnesses
mentally ill	(mentally ill)	mentally ill
mentally un	(mentally un!)	mentally unwell, mentally unstable, mentally unsound, mentally unhinged, mentally unfit

agoraphobia	(agoraphobi!)	agoraphobia, agoraphobic
anorexia	(anorexi!)	anorexia, anorexic, anorexia nervosa
anxiety	(anxiety)	anxiety
autism, autistic	(autism) or (austistic)	autism, autistic
binge eating disorder	(binge eating disorder)	binge eating disorder
bipolar	(bipolar!)	bipolar, bipolar disorder
body dismorphia	(body dismorph!)	body dismorphia, body dismorphic
borderline personality disorder	(borderline personalit!) or (bpd)	borderline personality, borderline personality disorder, bpd
bulimia	(bulimi!) or (bulimia)	bulimia, bulimic, bulimia nervosa
depression	(depress!)	depression, depressed, depressive, depressing
dissociative identity disorder	(dissociative disorder) or (dissociative identity disorder)	dissociative disorder, dissociative identity disorder
eating disorder	(eating disorder)	eating disorder
hypomania	(hypomania)	hypomania
hypermania	(hypermania)	hypomania
mania	(mania) or (mania!) or (manic) or (manic!)	mania, manic, maniacal, maniac, manic depression, manic depressive
multiple personality disorder	(multiple personality disorder) or (mpd)	multiple personality disorder, mpd
obsessive compulsive disorder	(obsessive compulsive disorder!) or (obsessive compulsive) or (ocd)	obsessive compulsive disorder, obsessive compulsive disorders, obsessive compulsive, ocd
paranoia	(paranoia)	paranoia
personality disorder	(personality disorder!)	personality disorder, personality disorders
postnatal depression	(postnatal depression)	postnatal depression
posttraumatic stress disorder	(posttraumatic stress) or (post traumatic stress) or (post-traumatic stress) or (ptsd)	posttraumatic stress, post traumatic stress, post-traumatic stress, ptsd
psychosis	(psychosis) or (psychotic)	psychosis, psychotic

schizophrenia	(schizophreni!) or (schizo)	schizophrenia, schizophrenic, schizo
seasonal affective disorder	(seasonal affective disorder!)	seasonal affective disorder, seasonal affective disorders
social phobia	(social phobia)	social phobia

TABLE 5.1. BREAKDOWN OF THE SEARCH TERMS.

Once the search terms had been collected and wildcards added, each term was joined using the OR Boolean operator denoting that the software should return articles that include any of the terms. This resulted in the final search term to be used in the LexisNexis database:

(mental illness!) OR (mental health!) OR (mental ill health) OR (mentally ill) OR (mentally un!) OR (agoraphobi!) OR (anorexi!) OR (anxiety) OR (autism) OR (autistic) OR (binge eating disorder) OR (bipolar!) OR (body dismorph!) or (borderline personalit!) OR (BPD) OR (bulimi!) OR (depress!) OR (dissociative disorder) OR (dissociative identity disorder) OR (eating disorder) OR (multiple personality disorder) OR (mpd) OR (obsessive compulsive disorder!) OR (obsessive compulsive) OR (ocd) OR (paranoia) OR (personality disorder!) OR (postnatal depression) OR (posttraumatic stress) OR (post traumatic stress) OR (post-traumatic stress) OR (ptsd) OR (psychosis) OR (psychotic) OR (schizophreni!) OR (seasonal affective disorder!) OR (social phobia) OR (bulimia) OR (hypomania) OR (hypermania) OR (mania) OR (mania!) OR (manic) OR (manic!)

FIGURE 5.2. SEARCH TERM FOR THE MI 1984-2014 CORPUS

In the next section, I discuss the rationale for the sampling from the dates covered.

5.3. Rationale for dates covered

In this section I discuss the rationale for the time period covered by the MI 1984-2014 corpus. My aim was to create a resource that represents the topic of mental health in the UK press and which is sizable enough for the findings generated by corpus analysis to be generalisable to the population from which the sample is drawn. For this reason, the dates covered had to be wide ranging enough to give adequate access

to the population I was trying to represent whilst also including any potential diachronic or synchronic changes.

In order to ascertain relevant periods of history to sample from, I researched the history of mental illness legislation. The most recent major mental health act was passed in November 1983 (subsequent amendments have been made since) and so I decided to start the data collection period in the January of 1983 to cover this period. In order to ensure that the data collected would be sizable enough, I chose to sample from 30 years of newspaper reports as this length of time is likely to be enough to cover changes in textual practices in newspaper reportage. As a result of this, the data is continuous and longitudinal, which facilitates diachronic analysis with a greater degree of nuance than other studies into newspaper reports on mental illness (for example, those that use discrete data to make claims about longitudinal trends in reporting, e.g. Goulden et al., 2007; Roberts et al., 2013). Due to the fact that the function of the corpus is to be broad in its coverage, it is also significantly larger than the corpora used in the few existing studies in linguistics that explore mental health reportage. For example, Atanasova et al. (2019) sampled from articles published between 2007 and 2015, creating a corpus of 485,186 words.

A further reason for selecting 30 years from 1983 was that the time period covered a relatively active area in legislation related to mental health. The time period covers The Mental Health Act 1983, The Mental Health (Patients in the Community) Act (1995), The Mental Capacity Act (2005), the amendments to the 1983 Mental Health Act in 2007 and the 2013 Mental Health (Discrimination) Act. For this reason, the dates chosen are broad enough to provide insight into any diachronic change but are also fitted to relevant time periods in the history of mental illness in a UK context. Moreover, collecting data from every week during the time period provides data that can be analysed for any synchronic changes, (e.g. how different newspapers report on the same story).

Some of the changes made as a result of changes in legislation on mental health and illness pertained to language and definitions of roles and mental illness terms.

Moreover, the reaction to some of these acts has received much press attention (particularly in relation to the move towards outpatient care). For this reason, the key changes stipulated by the Acts are detailed in Table 5.2.

Mental Health Act 1983	• First Mental Health Act since 1959 • Defined mental disorder as "mental illness, arrested or incomplete development of mind, psychopathic disorder and any other disorder or disability of mind" • Stipulated that a person with a mental disorder can be detained against their will for treatment (sectioning) if a professional deems this necessary
The Mental Health (Patients in the Community) Act 1995	• Provided provision for people with mental disorders to receive community care and community supervision
Mental Capacity Act 2005	• Listed the rights of patients who do not have the mental capacity to make decisions (including people with mental illness)
Amendments to Mental Health Act 1983 (2007)	• Change in definition of mental disorder from "any disorder or disability of the mind" • Introduced community treatment orders which mean that 'non-compliant' patients are treated in the community without their consent rather than be admitted back to hospital to be treated without their consent (move closer to care in the community model over inpatient care) • Widened definition of mental health professionals to include 'approved clinician' • Changed title of approved social worker to 'approved mental health professional' • Stated that Electroconvulsive therapy may not be used without consent from the patient
Mental Health (Discrimination) Act 2013	• Changed rules on whether members of parliament, jurors and company directors can serve in their relevant positions after being sectioned

TABLE 5.2. OVERVIEW OF MENTAL HEALTH LEGISLATION 1983-2013

The next section will discuss the software used to collect the data, LexisNexis.

5.3.1. LexisNexis

LexisNexis (https://www.lexisnexis.com/uk/legal/) is subscription-based software containing a database of legal and journalistic documents. Nexis is the part of this software that contains historic newspaper articles (from 1981 onwards) available for download in multiple file formats. The software allows researchers to set criteria for the articles it returns which means that it is routinely used in studies of this kind in both corpus linguistics and critical discourse analysis (see, for example, Baker et al, 2008; Baker & Levon, 2015; Gabrielatos & Baker, 2008; Grundmann & Krishnamurthy, 2010; inter alia). Nexis allows researchers to input their search terms and specify whether these should be present in the entire document, the headline of the article, the byline, etc. Furthermore, the researcher can stipulate both the date range of returns (e.g. last week, last month, custom range, etc.) and the source (e.g. UK Newspapers, UK Broadsheets, All English Language News, etc.). The researcher can select whether to group duplicate articles by high similarity and whether to include newswires, etc. The Nexis database of UK Newspapers contains broadsheets, tabloids and regional newspaper articles from the 1980s and is updated daily. A screenshot of the software is shown in Figure 5.3:

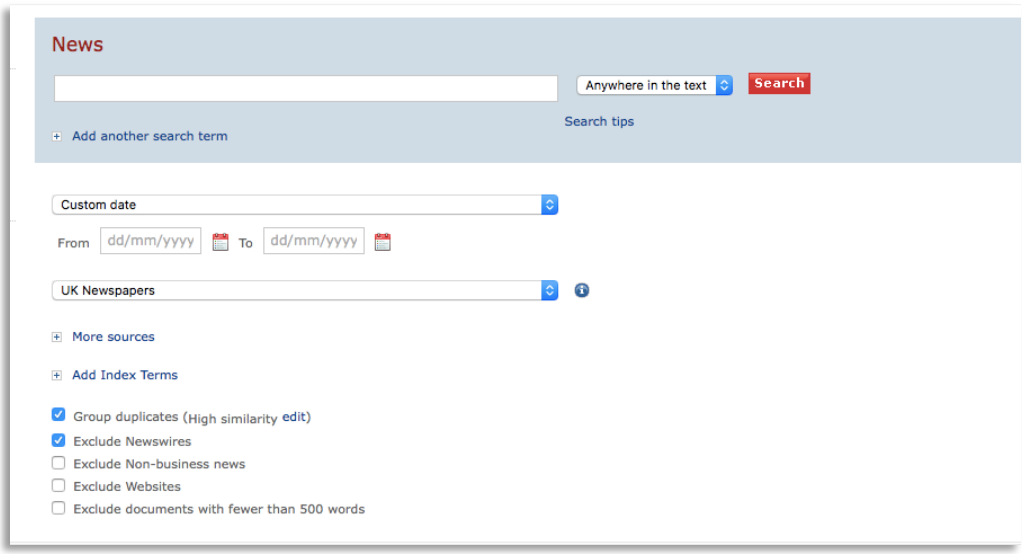


FIGURE 5.3. NEXIS SEARCH HOMEPAGE

Once the sampling frame has been selected, files can be downloaded from Nexis, in groups of 500 articles. The downloaded files also contain metadata listing the date and source of the publication, the byline of the article and the section of the newspaper that the article appeared in. An annotated screenshot of a Nexis file is shown in Figure 5.4.

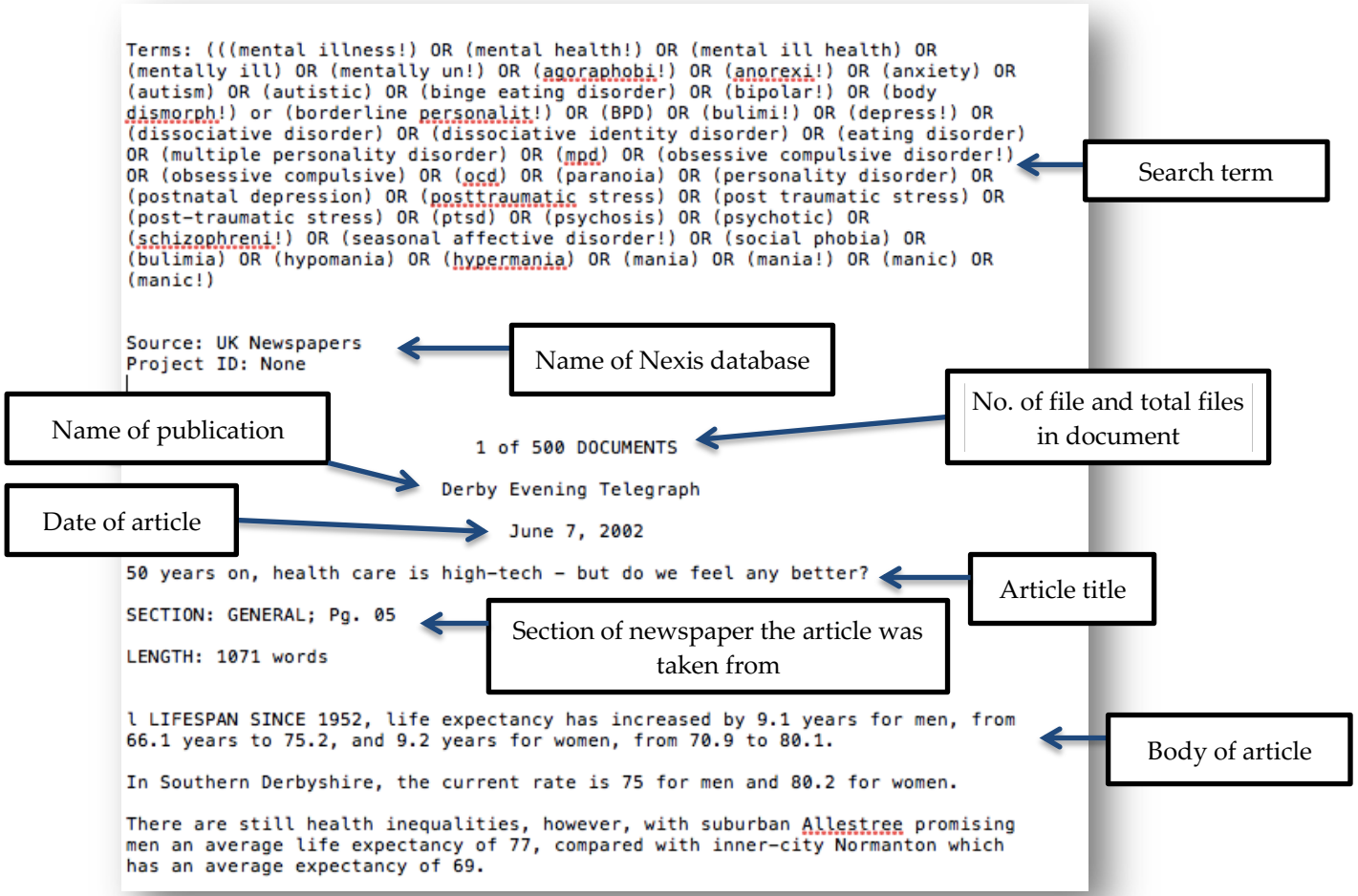


FIGURE 5.4. SCREENSHOT OF A NEXIS FILE AFTER DOWNLOAD

To the best of my knowledge, there is currently no way to automate the process of downloading files from Nexis, which poses problems for the researcher. The first of these is the issue of time. Downloading articles from Nexis often requires the researcher to download multiple files for just one category of interest, e.g. a week. For example, my sampling frame required me to search every seven days and often the number of articles returned required 3 or 4 separate downloads of 500 for this period. Each download takes around 5 minutes. Clearly, then, this is a manual and time-consuming process. Another problem is that Nexis does not allow the researcher to specify a naming strategy for the downloaded files. This means that after download, the files have to be renamed manually to preserve the chronological order.

5.3.2. Sampling frame for MI 1984-2014

The search terms and sampling frame were designed to be broad to allow for representative sampling and also to assess how much data was available for each period. The sampling frame I used in the data collection process was the search term detailed in the Section 5.2 within the Nexis UK Newspapers database with articles grouped by similarity and excluding newswires. The reason for choosing UK Newspapers over UK National Newspapers was that doing so allowed for the representation of both national and local news reports. This enabled me to identify local/national press journalistic differences as a variable should the need arise. Furthermore, as Atanasova et al. (2019) point out, sampling from local newspapers is a novel methodological decision as previous research has not focussed on investigating the variation between local depictions vs. national depictions of mental illness. Furthermore they note that

> Contrary to national media, local newspapers cater to smaller communities. It is, therefore, reasonable to assume that the readers and, indeed, writers of news articles published in local newspapers may personally know the individuals involved in the reported stories, resulting in more positive writing
>
> (Atanasova et al., 2019: 12)

Given that the final corpus may be used as resource to help answer a wide range of research questions and test hypotheses that exceed the scope of this thesis (for example, exploring the link between local newspapers and national newspapers), I wanted to make sure that these questions could be asked of the corpus at a later date should the need arise. I placed no restriction on where the search terms could appear (i.e. the search could return hits with just one instance of a word anywhere in the whole document). Given that the software restricts the amount of articles available for download, I searched for the terms of interest in 7-day periods[54] throughout the 30 years. Due to the sampling frame using a week as a sampling unit, some articles were collected from the first few days of January 2014. These articles were also collected.

As stated in Section 2.2.1, the Nexis database only contains newspaper articles from 1981 onwards. It also only holds certain newspapers in these years. For example, in the early 1980s the only newspaper stored in the Nexis database is *The Guardian*. For this reason, the number of hits returned from searches at the start of the 1980s was much lower than that of the end of the decade[55]. Files collected during the data collection process were named using this structure:

> YEAR_MONTH_DAY_YEAR_MONTH_DAY
>
> Example: 1998_09_21_1998_09_28
> (would contain articles published between 21 to 28 September 1998)

[54] When collecting data from LexisNexis in the early years of the search period, the search time was broadened as fewer hits were returned. For example, during 1984 it may have only been necessary to download files in a one month or two month period; however, in the later periods it was necessary to download the files by week.

[55] Indeed, the search results yielded no hits in 1983.

For 7 day periods that returned more than 500 hits and therefore required more than one file and additional specification, a number corresponding to the number of file was added to the file name:

YEAR_MONTH_DAY_YEAR_MONTH_DAY_NO. OF DOWNLOAD FOR THAT WEEK

Example: 1998_09_21_1998_09_28_01 → would contain files 1-499
1998_09_21_1998_09_28_02 → would contain files 500-999
1998_09_21_1998_09_28_03 → would contain files 1000-1499
Etc.

During the data collection process, all information about the files was added to a master spreadsheet for each year that detailed which week the data was from, how many hits the search had returned before similarity analysis, how many hits the search returned after similarity analysis and how the files were split if the 7-day period returned more than 500 articles.

1	DATE	HITS	HITS2	Files	SOURCES	NOTES
18	28/04/97	455	450			
19	05/05/97	435	425			
20	12/05/97	422	418			
21	19/05/97	403	403			
22	26/05/97	468	450			
23	02/06/97	444	442			
24	09/06/97	404	400			
25	16/06/97	428	425			
26	23/06/97	458	448			
27	30/06/97	491	484			
28	07/07/97	409	407			
29	14/07/97	450	434			
30	21/07/97	466	466			
31	28/07/97	481	463			
32	04/08/97	496	471			
33	11/08/97	596	574	1-499, 500-574		
34	18/08/97	540	540	1-499,500-540		
35	25/08/97	510	506	1-499, 500-506		
36	01/09/97	482	472			
37	08/09/97	501	475			
38	15/09/97	515	467			
39	22/09/97	582	492			
40	29/09/97	566	513	1-499, 500-513		
41	06/10/97	620	554	1-499, 500-554		
42	13/10/97	659	589	1-499,500-589		
43	20/10/97	641	578	1-499, 500-578		
44	27/10/97	611	548	1-499, 500-548		
45	03/11/97	574	511	1-499, 500_511		
46	10/11/97	586	517	1-499, 500-517		
47	17/11/97	627	559	1-499, 500-559		
48	24/11/97	739	656	1,499-500-656		
49	01/12/97	682	605	1-499, 500-605		
50	08/12/97	573	512	1-499, 500-512		
51	15/12/97	588	527	1-499, 500-527		
52	22/12/97	477	421			
53	29/12/97	574	521	1-499,500-574		

FIGURE 5.5. SCREENSHOT OF MASTER SPREADSHEET FOR 1997

The initial download yielded 3343 documents comprising c. 1,414,660 individual newspaper articles. Given that the articles range in length from relatively short articles (c. 600 words to over 7000 words) the number of words in this initial data collection was too large for any detailed analysis to be feasible, and moreover, would not be computable by mainstream corpus tools. In order for the data to be useable, it first needed to be sampled from and then cleaned. These issues will be discussed in more detail in the following section.

5.4. Data cleaning

One of the first ways to downsize the data was to rid it of any noise. Gabrielatos (2007) writes of noise in specialised corpora:

> there is a tension between, on the one hand, creating a corpus in which all the texts are relevant, but which does not contain all relevant texts available in the database, and, on the other, creating a corpus which does contain all available relevant texts, albeit at the expense of irrelevant texts also being included. Seen from a different perspective, the trade-off is between a corpus that can be deemed incomplete, and one which contains noise (i.e. irrelevant texts)
>
> (Gabrielatos 2007: 6)

In order to ensure that no highly relevant articles were removed, whilst also ensuring that noise was significantly reduced, a series of steps were designed in collaboration with Dr Hugo Sanjurjo-González, a programmer. These steps were then incorporated into a Python script. The first stage of the script was designed to do the following:

1. Split the files so that each individual article is in its own file
2. Delete duplicate files
3. Name the files to preserve the chronological order of the articles

4. Remove default metadata from the files (byline, author, newspaper section) but keep basic information about the article (title, date published)
5. Remove articles that do not contain at least three instances of a search term item
6. Remove any articles that contained the search term *depression* that also contained the words *economy*, *finance*, *money*

After this initial step, a wordlist was generated to assess any additional noise. At this point it became clear that some noise has been created as a result of the wildcards used on **mania* and *manic**. This resulted in a final cleaning step:

7. Delete articles that have been included as a result of **mania* and *manic** that do not pertain to mental illness (e.g. articles containing words such as *manicure* and *Romania*)[56]

The rationale for setting a cut off of three instances of a search term was that the data collection process had yielded a considerable amount of data to be analysed and had to be significantly reduced. Moreover, after initial assessment of the files returned in the data collection, it became clear that some regular features in the newspapers were creating a considerable amount of noise in the corpus as they did not relate to mental illness. For example, the 'Text Maniacs' feature in *The Sun*.

The procedure was successful in greatly reducing the number of tokens and increasing the overall relevance of the articles contained in the corpus. During this procedure, it also became apparent that due to the sampling frame, which searched for articles by week, the articles published in the first few days of January 2014 had been collected. Given that the earlier period of the sampling frame (1983 and up to July 1984) returned no or very few articles, the small number of articles collected from 2014 were included which resulted in the corpus representing articles published from

[56] Articles that contained one of these items (*Romania*, *manicure*) plus another search term, e.g. *bipolar disorder* were left in the corpus.

mid-1984 to early 2014. After the cleaning procedure was complete, additional file naming strategies were implemented. For example, the .txt file named '1986_01_01_1986_02_23_127.txt' relates to the 127th[57] article collected between the 1st January 1986 and the 23rd February 1986. The final streamlined corpus contains newspaper articles from 1st August 1984 to 5th January 2014 and comprises 64,521 articles and 50,972,932 words (calculated using AntConc-4).

5.5. Construction of subcorpora

Once the data had been cleaned, the next step involved the creation of subcorpora. I made the decision to create subcorpora for each year and for each illness. The reason for doing this is that creating subcorpora for each of the predictor variables (i.e. time and illness type) allowed for diachronic analysis (e.g. how each year differs in the reports on mental illness) as well as synchronic analysis (e.g. how news articles report on each illness type and how that compares with how other illnesses are reported). Furthermore, creating specific subcorpora based on the predictor variables allowed for more precise analysis of textual differences (or output variables) and how they pattern across each illness type, e.g. transitivity categories.

The construction of the subcorpora was again done in conjunction with a programmer. Based on the new naming schemes, which gave specific information about the day and year of publication, the year subcorpora (e.g. 1984, 1985, and so on) were created based solely on publication date. A python script written by the programmer was designed to read the year and date of publication of the article and add it to the relevant year subcorpus. The criteria for the illness subcorpora (e.g. AnxietyCorpus, BipolarDisorderCorpus) required more planning however. This was due to the fact that I wanted to avoid duplicating articles in the different illness

[57] This number relates to the position of the article in the original data collection period, not the final data contained in the corpus.

subcorpora. This meant that a very specific workflow had to be created to automate the process of adding files to the relevant subcorpus using some information contained in each individual article. The reason for not adding articles containing multiple terms to every relevant illness subcorpus was to allow for reliable results when making comparisons between the illness subcorpora. The first step in creating the illness subcorpora was to revisit the search terms and group them by illness. Wherever the search term was a general term, e.g. mental illness, and therefore did not pertain to a specific illness, it was added to a general mental illness subcorpus. The decisions in this step are shown below:

Subcorpus name	Terms to include
MentalIllness corpus	mental illness*, mental illness, mental ill health, mentally ill, mentally un*
Agoraphobia corpus	agoraphobi*
ASPD corpus	antisocial personality disorder, antisocial personality disorder*, aspd
Anorexia corpus	anorexi*
Anxiety corpus	anxiety
Autism corpus	autism, autistic
BingeEating corpus	binge eating disorder
BipolarDisorder corpus	bipolar*, bipolar, bipolar disorder, hypomania, hypermania, mania*, mania, manic*, manic, manic depressi*
BodyDismorphia corpus	body dismorph*
BPD corpus	borderline personalit*, borderline personality disorder, BPD
Bulimia corpus	bulimi*
Depression corpus	depressed, depression, depressive
DID corpus	dissociative disorder, dissociative identity disorder
EatingDisorder corpus	eating disorder, eating disorders*
MPD corpus	multiple personality disorder, multiple personality disorder*, mpd
OCD corpus	obsessive compulsive disorder*, obsessive compulsive, ocd
PostnatalDepression corpus	postnatal depression, postpartum depression, puerperal depression

PTSD corpus	posstraumatic stress, post traumatic stress, post-traumatic stress, ptsd
Psychosis corpus	psychosis, psychotic
Schizophrenia corpus	schizophreni!, schizo*,
SAD corpus	seasonal affective disorder*, seasonal affective disorder
SocialPhobia corpus	social phobia

TABLE 5.3. INITIAL LIST OF SUBCORPORA AND TERMS INCLUDED

It is important to state that at this stage, interpretation of the search terms was reintroduced again (albeit much more limited interpretation), as the categories required some conflation of terms. For example, by adding 'mania' and its related terms to the BipolarDisorderCorpus and not having it as a distinct illness it itself, I have imposed my view on the best way to group those illnesses. However, my rationale for my decision to do this is based on the observation that many articles featuring search terms related to mania were also often used in relation to Bipolar Disorder or symptoms of Bipolar Disorder, rather than as distinct illnesses.

In addition to the conflation of categories, some terms were added to enrich the terms added to the subcorpora. These were based on observations since the initial search terms were generated (for example, the use of 'puerperal depression' in the PostnatalDepression subcorpus). This again was based on the observation that this term was more frequent than postnatal depression in the data and, as a result, would be helpful in getting a representative sample of reports on postnatal depression. Once these terms were compiled, the first step in the workflow was to add the articles to the relevant subcorpus using the following decisions:

Decision	Example
If the article contains terms from only one subcorpus → add it to that subcorpus	An article contains only the words *autism* and *autistic* → it is added to the AutismCorpus
If the article contains terms from more than one subcorpus → add it to the corpus that the majority of the terms relate to	An article contains 5 instances of *depression* and 3 instances of *agoraphobi** → add to the DepressionCorpus

TABLE 5.4. INITIAL WORKFLOW DECISIONS FOR ILLNESS SUBCORPORA CONSTRUCTION

Once the script had been developed, written, and run, it became apparent that there were too many conflicts in the script (where an equal number of terms occurred in one article, each pertaining to a different subcorpus) for the current design of the subcorpora to be feasible. The reason for this was that it was not simply the case that articles reported on one illness and one illness only, and, as stated in the discussion of search terms earlier, many illnesses are related, i.e. in their symptoms. As a result, it was often the case that an article reported on mental illness generally, making reference to multiple illnesses in the article. In cases like this, there had to be a decision made (and a decision that was possible to automate) for which subcorpus to attribute the article to. With this in mind, I decided to conflate some of the illness subcorpora further. First, the bulimia, anorexia, eating disorder and binge-eating subcorpora were merged to create one EatingDisorderCorpus[58]. Second, I also decided to conflate the multiple personality disorder (MPD) and borderline personality disorder (BPD) subcorpora to create one PersonalityDisorderCorpus. The rationale for this was that the illness subcorpora are to be used for cross comparison only and therefore specific differences between the terms are not pertinent to the analysis. Moreover all these terms would be categorised as eating disorders or personality disorders in DSM-V anyway and therefore the terms are still representative. Once this was done, the number of conflicts in the script was greatly reduced. The second decision to limit

[58] This is an interesting initial insight into the corpus as the number of conflicts indicates that the terms are often reported together and therefore related linguistically to a certain extent.

common conflicts in the script was to set a rule that stated that any articles containing 'mania' and 'depression' (or some variety of those lemmas) should not be attributed to the DepressionCorpus, but rather the BipolarCorpus. The reason for this was based on the fact that these two words pertain to the symptoms of bipolar disorder specifically, and therefore it is sensible to assume that the article as a whole discusses bipolar disorder. Despite the number of conflicts being greatly reduced by this stage, there was still an issue with terms occurring equally often in some articles. To remedy this, I decided to prioritise some subcorpora over others. This resulted in a third step to the workflow:

Decision	Examples
If the article contains an equal number of search terms from two or more corpora → then prioritise the subcorpora in the following way (in descending order of priority) (All other subcorpora) AnxietyCorpus DepressionCorpus MentalIllnessCorpus AutismCorpus 	An article contains 4 instances of *depression*, 4 instances of *autism* and 3 instances of bipolar disorder → it is added to the DepressionCorpus An article contains 4 instances of anorexia and 4 instances of mental illness → it is added to the EatingDisorderCorpus An article contains 3 instances of mental illness, 3 instances of depression and 3 instances of autistic → it is added to the DepressionCorpus

TABLE 5.5. WORKFLOW DECISIONS FOR ILLNESS SUBCORPORA CONSTRUCTION (THIRD PHASE)

Low priority status was only given to four subcorpora in order to reduce the number of conflicts. The reason for giving low priority to the AutismCorpus was that autism was not a key area of study in this analysis; rather it was included as a search term to identify articles that may not have been otherwise accessed by LexisNexis, for this reason it was an obvious decision to prioritise the other subcorpora over the AutismCorpus. The reason for giving low priority to the MentalIllnessCorpus was that it was a designed to catch any articles that were not covered by specific illnesses;

as such, the MentalIllnessCorpus was another obvious candidate to reduce the number of conflicts. The reason for giving low priority to the DepressionCorpus and the AnxietyCorpus was that depression and anxiety are symptoms of many other illnesses and as a result the articles containing words pertaining to depression and anxiety may have been reporting on another, more specific illness. In addition to this, ranking depression lower than anxiety was a practical decision. This was done to limit the number of articles added to the depression corpus automatically that did not directly report on depression but were added to the depression corpus because there were more instances of 'depression', 'depressed', 'depressive' than the specific illness in the article. These decisions greatly reduced the number of conflicts further. After this stage, the conflicts in the script were minimal (<20). This then meant that the remainder of the items could be manually added to the relevant corpora. Using the decisions described above, the workflow can be summarised (with examples) as follows:

Workflow:

I. Check for most frequent term in article
II. If there are two terms then:
    a. If one term belongs to BipolarDisorderCorpus and the other to DepressionCorpus, → include term into BipolarDisorderCorpus
    b. If it is not the case then discard MentalIllnessCorpus terms if any
    c. If there is still more than one term then discard DepressionCorpus terms if any
    d. If there is still more than one term then discard AnxietyCorpus terms if any
III. If there are more than two:
    a. If the terms occur equally often and belong to the same corpus, the document must be included in that corpus.
    b. If there are the same quantity of terms from two or more corpora then:
        i. Discard terms from MentalIllness_corpus if any
        ii. If there are still terms from more than one subcorpus then discard DepressionCorpus terms if any
        iii. If there are still terms from more than one subcorpus then discard AnxietyCorpus terms if any

FIGURE 5.6. SUMMARY OF WORKFLOW DECISIONS FOR ILLNESS SUBCORPORA CONSTRUCTION

An overview of the illness subcorpora, including the revised version of the illness subcorpora, terms included and number of articles in each illness subcorpus is detailed in Table 5.6. The subcorpora for ASPD, body dismorphia and postnatal depression are not present because no articles reporting on those illnesses and those illnesses only were present in the data collected (instead they will have been attributed to the relevant corpora due to the decision tree in the cleaning process.) In addition to the Table 4 that provides an overview of the illness subcorpora, Table 5.7 provides an overview of the year subcorpora.

Illness subcorpus	Terms included	No. of articles	Average article length	No. of tokens	Percentage of total
Agoraphobia	agoraphobi*	198	875	173,292	0.35%
Anxiety	anxiety	3572	924	3,301,718	6.63%
Autism	autism, autistic	11209	549	6,153,039	12.36%
Bipolar	bipolar*, bipolar, bipolar disorder, hypomania, hypermania, mania*, mania, manic*, manic, manic depressi*	624	709	442,325	0.89%
DID	dissociative disorder, dissociative identity disorder	7	1669	11,685	0.02%
Depression	depressed, depression, depressive	33628	831	27,937,710	56.11%
Eating Disorder	anorexi*, binge eating disorder, bulimi*, eating disorder, eating disorders	5781	782	4,523,057	9.08%
Mental Illness	mental illness*, mental illness, mental ill health, mentally ill, mentally un*	3066	670	2,053,493	4.12%
OCD	obsessive compulsive disorder*, obsessive compulsive, ocd	644	784	505,034	1.01%
PTSD	posstraumatic stress, post traumatic stress, post-traumatic stress, ptsd	1166	805	938,521	1.89%
Personality Disorder	multiple personality disorder, multiple personality	1171	936	1,096,012	2.20%

	disorder*, mpd, borderline personalit*, borderline personality disorder, bpd				
Psychosis	psychosis, psychotic	854	769	656,321	1.32%
SAD	seasonal affective disorder*, seasonal affective disorder	23	860	19,775	0.04%
Schizophrenia	schizophreni!, schizo*,	2542	763	1,940,789	3.90%
Social Phobia	social phobia	44	798	35,118	0.07%

TABLE 5.6. OVERVIEW OF THE ILLNESS SUBCORPORA

Illness subcorpus	No. of articles	No. of tokens	Average article length	Percentage of total
Agoraphobia	198	173,292	875	0.35%
Anxiety	3572	3,301,718	924	6.63%
Autism	11209	6,153,039	549	12.36%
Bipolar	624	442,325	709	0.89%
DID	7	11,685	1669	0.02%
Depression	33628	27,937,710	831	56.11%
Eating Disorder	5781	4,523,057	782	9.08%
Mental Illness	3066	2,053,493	670	4.12%
OCD	644	505,034	784	1.01%
PTSD	1166	938,521	805	1.89%
Personality Disorder	1171	1,096,012	936	2.20%
Psychosis	854	656,321	769	1.32%
SAD	23	19,775	860	0.04%
Schizophrenia	2542	1,940,789	763	3.90%
Social Phobia	44	35,118	798	0.07%

TABLE 5.7. OVERVIEW OF THE YEAR SUBCORPORA

The next section will discuss some of the practical issues of constructing a corpus and reflect on this process.

5.6. Practical issues: linguistics, programming and the nature of language

The process of dealing with huge quantities of electronic language data brings to light one of the issues for researchers using corpora, which is that in many ways the expertise needed to build corpora and conduct corpus analysis crosses (inter)disciplinary boundaries. Research of this kind necessitates knowledge of programming languages, regular expressions, how to exploit existing databases, how to use existing corpus software and understand the programming decisions built into them, how to carry out statistical tests and which statistical tests to use for which analytical method[59], and practical issues such as having the facility to store the sheer amount of data needed to build a corpus. All of this is before an in-depth knowledge of the theoretical and methodological underpinnings of any linguistic analysis.[60] The interdisciplinary nature of corpus linguistics (particularly the link between corpus linguistics and programming) is a topic that has received some attention over the past few years in particular. One notable example was a British Association for Applied Linguistics (BAAL) special interest group meeting which was centred around the question "Does a corpus linguist need to be a computer programmer?"[61]. It is evident from the processes I have outlined here that had I not been fortunate enough to work with a programmer, the final corpus would have been a much lesser product, or would have taken an unfeasible amount of time to compile. For instance, without the advantages of automation, it would have been necessary to carry out each task manually, and this would have resulted in a corpus that covered a considerably narrower selection of years and illness types.

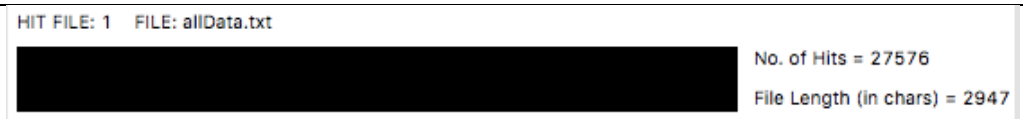
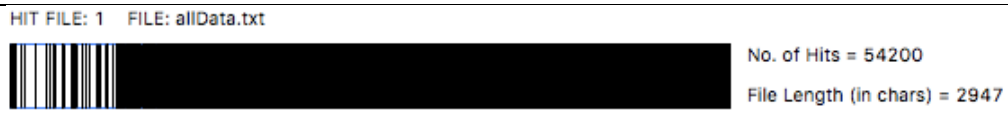
[59] This is a useful skill to have simply to defend the statistical tests chosen to other researchers.
[60] When confronted with this task, a researcher can take solace in the fact that the difficulty of needing to wear numerous analytical 'hats' has long been documented. In his essay on statistics and style, Bailey (1969; cited in Stubbs 1994: 216) states that "the history of statistics and style shows few cases in which genuine expertise in language, literature and statistics have been combined in one investigator."
[61] BAAL Corpus Symposium, held at Aston University, May 6, 2016.

However, there were some steps in the process of creating the illness subcorpora where I felt there to be a tension between what was needed to make the project sensible from a programming point of view and what I felt was important as a linguist. An example of this was the prioritising of one illness subcorpus over another. My feeling on this was that it imposed a hierarchy on the significance of particular illnesses which I felt was not how language works. That is, the tendency in corpus linguistics is to privilege the word over the text (criticisms of this tendency in corpus linguistics have been made by Egbert & Schnur (2018) in relation to keyness analysis, for example) and production over comprehension. To give an example of what I mean by this, consider an article that features terms relating to a range of illnesses. Despite this, it may be the case that on reading the article we are clear that the text overall relates to schizophrenia. In taking frequency of terms as the basis for decisions about which subcorpus to add the text to, we have to prioritise the frequency of individual words over textual coherence. In creating the subcorpora using the computational methods I did, I have to rely on the belief that, at least to a certain extent, the topic of a text is indicated by how frequently a word appears in a text (i.e. generally speaking, the more a term from a subcorpus appears, the more likely it is that the text is reporting on the relevant illness).

The fundamental tension caused by having to make these decisions is that, as a linguist, my belief is that language is produced by individuals and therefore the texts may exhibit nuanced differences. However, the greater the number of automated decisions (i.e. to which subcorpus an article is added), the more it is necessary for the linguist to treat the text as being part of a homogenous group (resulting in the possibility that these nuanced differences become harder to identify). This is not to say that creating subcorpora using automatic methods should be avoided; because using programming opens up many avenues of exploration to the linguist.

5.7. The MI 1984-2014 Corpus: assessing relevance and distribution of terms

In this section, I will provide a brief overview of the final data to demonstrate that the terms searched for are present and evenly distributed across the MI 1984-2014 corpus. I will show this using concordance plots. In order to view the distribution of terms across the whole corpus, I used the Bash[62] concatenate command to merge the year subcorpora into a single file. As a result of this, the occurrence of the terms should be in rough date order, where the leftmost section of the concordance plot represents 1984 and the rightmost represents 2014. The box represents the corpus; the vertical lines represent instances of terms caught by the lemma across the sample. Whilst concordance plots are limited in what they can show (for example they do not show all the relevant words for each illness type), and they do not show the dispersion in great detail, the plots do show that the terms searched for appear to be well represented and evenly dispersed in the corpus. This suggests that the cleaning process has been successful in eliminating unnecessary noise from the corpus. Table 5.8 shows the concordance plots for each lemma searched.

Lemma	Concordance Plot Generated using AntConc (Anthony, 2017)
*agoraphobi**	
anxiety	
autism	

[62] Bash is a command language. It is the default command line interface on Unix-based systems such as macOS.

bipolar	HIT FILE: 1 FILE: allData.txt No. of Hits = 5378 File Length (in chars) = 2947
disassociative identity	HIT FILE: 1 FILE: allData.txt No. of Hits = 31 File Length (in chars) = 2947
depression	HIT FILE: 1 FILE: allData.txt No. of Hits = 91515 File Length (in chars) = 2947
eating disorder	HIT FILE: 1 FILE: allData.txt No. of Hits = 7877 File Length (in chars) = 2947
mental health	HIT FILE: 1 FILE: allData.txt No. of Hits = 34805 File Length (in chars) = 2947
mental illness	HIT FILE: 1 FILE: allData.txt No. of Hits = 16297 File Length (in chars) = 2947
obsessive compulsive	HIT FILE: 1 FILE: allData.txt No. of Hits = 1727 File Length (in chars) = 2947
ptsd	HIT FILE: 1 FILE: allData.txt No. of Hits = 4507 File Length (in chars) = 2947
personality disorder	HIT FILE: 1 FILE: allData.txt No. of Hits = 2177 File Length (in chars) = 2947
psychosis	HIT FILE: 1 FILE: allData.txt No. of Hits = 3248 File Length (in chars) = 2947
seasonal affective	HIT FILE: 1 FILE: allData.txt No. of Hits = 883 File Length (in chars) = 2947

*schizophreni**	HIT FILE: 1 FILE: allData.txt No. of Hits = 15159 File Length (in chars) = 2947
social phobia	HIT FILE: 1 FILE: allData.txt No. of Hits = 442 File Length (in chars) = 2947

TABLE 5.8. CONCORDANCE PLOTS SHOWING A SAMPLE OF TERMS IN THE MENTAL ILLNESS 1984-2014 CORPUS

In addition to this, a keyword comparison of the data sample and the SiBol English Broadsheet Newspapers 1993-2013 corpus[63], conducted using Sketch Engine (Kilgariff et al., 2004), shows that the top 50 keywords are what one would expect to see in a corpus of mental illness discourse. Moreover, the keywords are relevant even in comparison to another specialised corpus of newspaper data. These top keywords suggest that the data sample is representative of the population targeted (newspaper representations of mental health) because they are overrepresented in the MI 1984-2014 corpus compared with the SiBol corpus, and is further evidence that unnecessary noise in the corpus that would affect the analysis has been sufficiently minimized (aside from BODY, TITLE and DATE which refer to the metadata that was purposely kept in the files). This suggests that the final corpus is representative of the population and, as a result, generalisations to follow from the analysis will be reliable. Table 5.9 shows the top 40 keywords in the corpus.

[63] The SiBol English Broadsheet Newspapers 1993-2013 corpus comprises 650 million words of English Broadsheet Newspapers. The corpus was compiled by research teams at the Universities of Siena and Bologna and is available on Sketch Engine.

Rank	Term	Keyness Score	Freq	Ref Freq	Rank	Term	Keyness Score	Freq	Ref Freq
1	depression	3.72	83676	14363	21	depressed	1.8	24764	9122
2	BODY	2.97	57812	597	22	anxiety	1.79	25008	13023
3	DATE	2.97	57716	442	23	problems	1.73	37536	123916
4	TITLE	2.97	57706	445	24	Dr	1.7	35658	117536
5	mental	2.72	56460	28731	25	treatment	1.68	29426	75209
6	autism	2.46	43320	3290	26	feel	1.67	41785	174739
7	she	2.41	216741	957729	27	drugs	1.64	25285	52173
8	I	2.28	646631	3511339	28	disorders	1.64	19300	4969
9	her	2.23	244554	1227478	29	condition	1.63	24441	47762
10	health	2.1	57028	155325	30	patients	1.62	26386	65995
11	illness	2.08	34702	18872	31	parents	1.59	32676	127019
12	She	2.04	80769	324171	32	suffering	1.57	21269	37251
13	my	2.02	145281	750596	33	symptoms	1.57	18242	13234
14	children	2.01	74375	294050	34	autistic	1.56	16724	1663
15	me	1.95	107377	536782	35	stress	1.54	18889	25659
16	help	1.87	62849	262574	36	says	1.52	82859	585562
17	eating	1.84	28930	29974	37	MMR	1.51	15155	1306
18	disorder	1.83	25743	10387	38	brain	1.51	18751	32821
19	people	1.82	141883	847693	39	child	1.51	29004	122503
20	life	1.81	74135	364707	40	suicide	1.5	18411	33827

TABLE 5.9. TOP 40 KEYWORDS IN MI 1984-2014 CORPUS COMPARED WITH SIBOL ENGLISH BROADSHEET NEWSPAPERS 1993-2013.

5.8. Caveats of the corpus construction procedure

Before moving on to the conclusion section of this chapter it is first important to state the methodological caveats identified and offer responses to them. The first caveat is that the section of the newspaper the article appears in (e.g. editorial, opinion piece) is unaccounted for in the final corpus (i.e. this information was removed prior to uploading the corpus to the various corpus tools). The section of the newspaper that the article appears in may be meaningful in determining variation across it. However, the decision to remove information from the articles pertaining to newspaper section was based on the findings reported by Ohlssen (2017: 301), who found that mental health was talked about in "practically all the newspaper sections". Furthermore, this information, although not contained in the files included in the final corpus, is available in the raw files should this variable need to be explored in more detail.

A second caveat relates not to the design of the corpus, but rather the nature of newspaper data generally and newspaper data downloaded from LexisNexis specifically. Unlike newspaper articles generally, that are heavily multimodal, the newspaper articles contained in a corpus have to be stripped of any images or multimodal aspects relating to the size and colour of fonts in order to be read by the software. As a result of this, it is the case that some meaningful elements of the texts will be lost. This is important to state because prior research into the representation of mental illness suggests that images do play a role in creating stigma (Angermeyer et al., 2005). However, this is an unavoidable issue related to the methods adopted in corpus linguistics, although recent developments in multimodal corpus tools promise exciting developments in this area (see Knight, 2011). Another caveat related to data sampled from local and national newspapers is the issue of article duplication. As outlined in Section 5.4, exact duplicates have been removed from the corpus; however, because some national newspapers have regional editions (e.g. *The Daily Telegraph* Scotland Edition) that may feature very similar, although not identical, articles, the

possibility that some minor duplication may occur in the corpus is possible. However, this noise has been greatly reduced by removing exact duplicates. Moreover, one could take the view that duplication is necessary in order to accurately represent the population (e.g. all UK reporting on mental health and illness in this period).

5.9. Conclusion

In this chapter I have described the process and rationale for building the Mental Illness 1984-2014 corpus, focussing specifically on the compilation of search terms. I made a case for why researchers working in corpus linguistics should pay greater attention to their search terms when compiling corpora but also when analysing existing corpora, as the terms used are in themselves a reflection of the (compiler's view of the) target population. In this chapter I have also introduced the Mental Illness 1984-2014 corpus and demonstrated that the methods underlying its compilation have ensured that it is representative, that the terms targeted are well distributed and, as a result of this, that the results stemming from an analysis of it will be generalizable beyond the corpus.

In sections 5.2-5.3, I described the rationale for the sampling frame used in the data collection process, showing how the time period and search terms selected were relevant and fitted to the time, showing that the period was one in which major legislation changes in mental health and illness happened. I also discussed how the sampling frame used was wide enough to facilitate the analysis of synchronic and diachronic aspects of the texts, and outlined how the corpus and subcorpora have been designed to make the comparison of the years and different illness types straightforward. Moreover, I have exemplified how this method of compiling the data means that a significantly longer time period is covered, resulting in significantly more data being available for analysis than is used in other studies in linguistics exploring mental illness reportage to date. I have also shown how the data in this thesis is continuous and therefore better represents mental illness discourse than

previous longitudinal studies in research into mental illness in the press. Taken together, these decisions demonstrate that the Mental Illness 1984-2014 corpus is a representative corpus that can be used to answer a wide range of research questions, even those that fall outside the scope of this thesis. As a result, the creation of this resource constitutes one of the significant innovations of this thesis.

Sections 5.4 and 5.5 outlined the corpus cleaning procedures and the decision making process for the creation of subcorpora. By providing transparent information about the decisions made during the corpus construction process, I have shown that the method is replicable. Additionally, section 5.6 offered a reflective account of the practical issues associated with combining perspectives from linguistics and computer programming. I showed how the terms from the search term were distributed in the MI 1984-2014 corpus in Section 5.7. In Section 5.8, I outlined and responded to any methodological caveats.

In the next chapter, which is the first analysis chapter in this thesis, I use the MI 1984-2014 corpus to explore the terms 'mental illness' and 'mental health'.

6. The shifting meaning of mental health and mental illness

In this chapter I address the following research question listed in the introduction "How are the terms 'mental illness' and 'mental health' used in the MI 1984-2014 corpus?". I argue in this chapter that the meaning of 'mental health' and 'mental illness' has changed over the time period sampled (1984-2014). Furthermore, I argue that the change in the meanings of these concepts is consistent with pragmatic accounts of language change whereby change is socially-motivated (Ariel, 2008; Clark, 2016; Traugott & Dasher, 2002). Specifically, I argue that the changes in meaning I have identified are in keeping with accounts of language change that focus on euphemistic usages of a lexical item. In this chapter then, I explore the labels associated with the topic of this thesis: 'mental illness'[64] and 'mental health'. The reason for exploring the labels associated with these concepts is that no study to date has explored diachronic change in these terms which suggests that researchers have presumed the meaning associated with these labels to have been fixed over time. As a result of taking these labels for granted (i.e. presuming their meaning is fixed over time) no study to date has explored the diachronic change in these labels, particularly in studies reporting on mental illness in the press. Moreover, my initial observation from the MI 1984-2014 corpus was that 'mental health' and 'mental illness' are used interchangeably and therefore the notion that these labels were being used as near synonyms provided a hypothesis (i.e. 'mental illness' and 'mental health' are used as near-synonyms) that could be tested using real-world language data.

Examining in more detail how these terms are used makes it possible to provide a usage-based definition of the two terms. Moreover, how the concepts associated with these two terms are discursively constructed could have a bearing on the possible participants and processes that are also discursively constructed. For example, if

[64] I use italics to denote a concept, e.g. the concept of *mental health*, and scare quotes to denote the linguistic form used to refer to a concept, e.g. 'mental health' can be used to refer to the concept of *mental illness.*

mental health is conceptualised as being in a state of *mental wellness*, i.e. not *illness*, then one might expect to find fewer references to medical experts in newspaper reports on mental health due to the fact that mental health (relating to wellness) is not a pathology, and therefore may not warrant medical expert spokespeople. In contrast, where 'mental illness' is referred to, we might expect to see a greater number of medical experts being referred to, because *mental illness* is pathologised and therefore expert voices on this topic may offer insight into, for example, the diagnosis and treatment of a mental illness. Furthermore, in articles where 'mental illness' is referred to, we might expect to have more instances of processes such as 'diagnose' or 'treat' in contrast to articles discussing *mental health*, which may place greater emphasis on 'maintaining' or 'supporting'. The way that participants are named and what processes are represented in the data will be explored in more detail in Chapters 7 and 8 respectively, so it is important to outline before then what the concepts associated with 'mental health' and 'mental illness' are.

As previously stated, despite the existence of many studies of mental illness and health in the press (e.g. Sapogna et al., 2017; Søgaard et al., 1995; Stuart, 2003; Whitley & Wang, 2017), no study to date has explored the semantic content of these terms. This raises questions about the results of studies into the representation of mental health and illness in the press because the uncritical use of the terms 'mental illness' and 'mental health' means that any findings are predicated on the assumption that the users of the language being analysed conceptualise 'mental health' and/or 'mental illness' in the same way that the researcher does (which is unlikely given that the majority of existing research into this area has been conducted by psychiatrists, or specialists in the field). Furthermore, some of the existing research is diachronic (Goulden et al, 2011; Roberts et al., 2013; Whitley & Berry, 2013) and rests on the assumption that the concepts associated with the labels 'mental illness' and 'mental health' have stayed static across the time periods in question. It could be argued that differences in the concepts associated with the two labels are irrelevant because the researcher is only interested in the discourse surrounding the labels, and not the labels

themselves. However, diachronic analyses that treat 'mental health' and 'mental illness' as having fixed meanings across the time period leave no space for observing subtle changes in the discourse across time, which may have implications for any possible findings and for advancing the field, e.g. by identifying more specific research questions. For example, Whitley and Berry (2013) explore the representation of mental illness in Canadian print media between 2005-2010 using 'mental illness' and 'mental health' as a search terms. They adopt content analysis as their methodological approach citing that such an analytical approach ensures "consistency, validity, and reliability." (2013: 109). In their research, they do not query any change in the terms 'mental health' or 'mental illness' during the time period, neither do they offer any definition of 'mental illness' or 'mental health' to provide insight into how these two terms are related or how they differ. Whitley & Berry's (2013) research aims to analyse whether there has been a change over the time period in whether the newspaper reports they collected portray mental illness negatively using a coding scheme that centred around three themes: violence, criminality and danger. Whitley & Berry (2013) state explicitly how their coding scheme and methodological approach were controlled to allow for consistency and reliability, yet they appear to overlook two key variables – whether the meaning of 'mental health' and 'mental illness' differed across the time period. Further evidence that the authors overlooked any diachronic change in the labels 'mental health' and 'mental illness' is that throughout the research, the authors use 'mental health' and 'mental illness' interchangeably, which suggests that they view the two concepts as closely related or synonymous. In treating 'mental health' and 'mental illness' as near synonyms with fixed meanings, a vast amount of potential variation has been neglected, e.g. whether and when new terms are added to the discourse of mental illness. For the non-linguist, new terms may not be an area of interest, however, I argue that if the discourses around mental illness are of interest then new terms should be too, as new terms are revealing of how the existing terms are conceptualised, e.g. why would a new term be introduced to the discourse of mental illness if the existing terms were sufficient to

convey the meaning a person wants to communicate? Exploring possible language change in the labels 'mental illness' and 'mental health', then, provides insight into societal views on mental illness.

Moreover, if we look closer at 'mental illness' and 'mental health' in use, it is possible to see clear distinguishing features between the two terms. The findings reported in this section then, contribute to our understanding of the concepts of *mental health* and *mental illness* in a British English context, but also provides evidence to show that research that uses search terms to collect newspaper articles on mental health and illness should be mindful of variation in the semantic content of the terms searched for during the data collection procedure.

To illustrate this last point, I will refer to the existing literature in psychiatry, as this is the discipline in which most research into press representations of mental health have been conducted. A well-cited publication by Wahl et al. (2002) into press representations of mental illness in the US refers to prior research conducted by Day & Page (1986) in a Canadian context. Wahl et al. (2002) argue that few studies like that of Day & Page (1986) have been conducted in a US context, and this is how they set their own work in context. What we can presume then, is that the phenomenon that Day & Page (1986) analyse and the phenomenon that Wahl et al. (2002) analyse is the same and only the context changes (e.g. geographic location). However, if one looks in more detail at the methods sections of the two publications there is a difference. Day & Page (1986) construct their collection of newspaper articles based on the newspaper index term 'mental health' and Wahl et al. (2002) collect theirs using the search term 'mental illness'. No attention is paid to this difference, but as I show in Section 6.1 such differences in labels could potentially bias analysis.

In this chapter, I demonstrate how the meanings of terms contained in the semantic field of *mental health* and *mental illness* have shifted using evidence from language in use. Through my analysis, I argue that linguistic analysis provides researchers working in mental illness studies with insight into how mental illness is perceived in society. I argue that this insight is more robust than static and dated

dictionary definitions, making the case that linguistic analysis provides a means of tracking emergent semantic change.

6.1. Exploring *mental illness* and *mental health* in the MI 1984-2014 corpus

The first step in ascertaining whether there is a difference between the labels 'mental illness' and 'mental health' was to establish their usage across the time period. Figure 6.1 shows the relative frequency per million words (hereafter pmw) of the labels across the year subcorpora. In order to remove instances where 'mental health' formed part of a bigger phrase, i.e. 'mental health act', 'mental health legislation', relative frequencies were only taken for instances of 'mental health' that did not occur within 5 words to the left or right of 'act' or 'legislation'. Linear trend lines have been added for both terms to show the overall trend for each term more clearly.

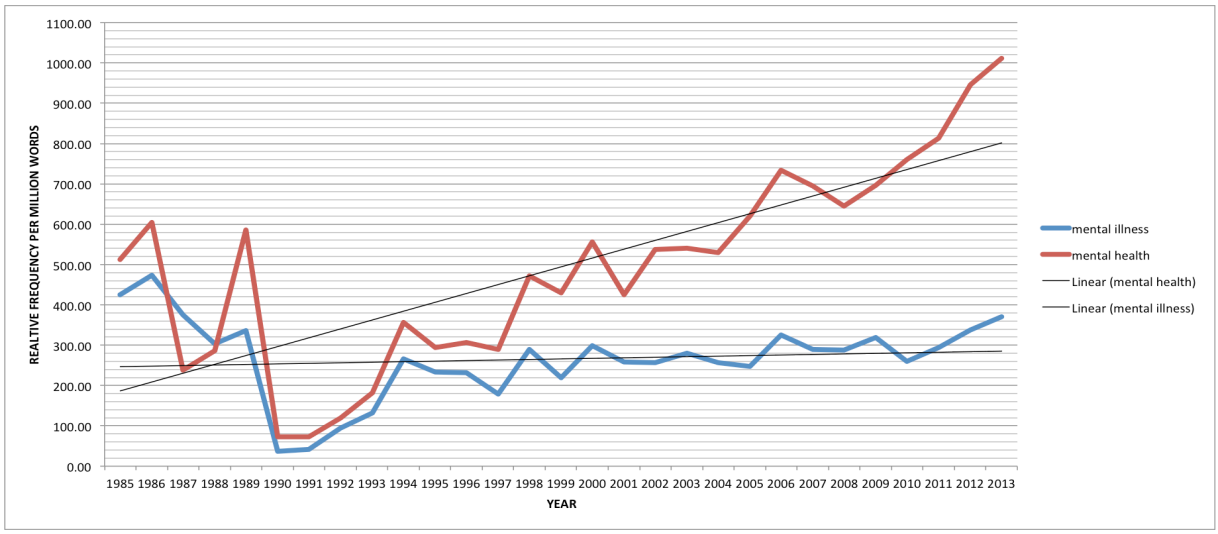


FIGURE 6.1. FREQUENCY OF 'MENTAL ILLNESS' AND 'MENTAL HEALTH' ACROSS THE YEAR SUBCORPORA

Figure 6.1 shows that both terms increase in use over the time period, however the increase in 'mental health' is significantly higher than that of 'mental illness'. The graph also shows that the usage of both terms rise and fall in the same years up to c.2008. In addition, exploring the shape of the overall trend of the two terms reveals that the pattern for both the terms over the time period is the same. This indicates that

periods of increased usage of 'mental illness' and 'mental health' are correlated, i.e. increased usage of 'mental health' correlates with increased usage of 'mental illness'. The fact the two terms are positively correlated indicates that the two terms are closely related within a semantic field. Moreover, the increase in both terms over the time period is indicative that the number of articles on the topic of mental health and mental illness are increasing overall, which would support previous findings that mental illness is a topical, and popular press issue (Ohlsson, 2017). By way of attempting to contextualise the fact that usages of 'mental health' are more frequent and are rising more substantially than 'mental illness', one may hypothesise that *mental health (*i.e. the full range of mental states – illness and wellness) is more prevalent societal issue than it was previously which has resulted in a greater number of articles on this issue. However this is not the case. The increase in 'mental health' usage is not due to an increase in articles that report on a range of mental health states (including *good mental health*) in society, but rather an increase in articles reporting on *mental illness.* I show my linguistic evidence for this claim throughout this chapter, but my interpretation is attested by statistics on mental illness in the UK. For example, the number of detentions[65] under the mental health act (MHA) has increased year on year since the Care Quality Commission started measuring the use of the MHA in 2009. Furthermore, the proportion of the English population with a mental disorder increased from 15.5% in 1993 to 17.6% in 2007. What the increase in 'mental health' suggests in light of these statistics showing that mental illness is rising is that there could be a developing preference to refer to all phenomena related to *mental health* (including *mental illness*) as 'mental health'. What the increase in 'mental health' over 'mental illness' suggests is that 'mental health' is potentially being used as a euphemistic term for the illness dimension of *mental health* to avoid the discussion of an 'emotionally marked domain' which is considered taboo (Blank, 1999). Burridge (2012: 67) identifies both *madness* and *disease* as taboo subjects that give rise to

[65] Detention here refers to the forced hospitalisation of someone where the hospitalisation was due to an enforcement of the Mental Health Act.

euphemistic forms, stating that "Since the 1980s, gender, sexuality, disability and race have become so highly-charged that speakers will shun anything that may be interpreted as discriminatory or pejorative" (Burridge, 2012: 67). It is possibly the case then, that the more general term 'mental health', is being used as a means of "obscuring and disguising disagreeable reality." (Burridge, 2012: 66). Furthermore, the use of a general term to refer to a specific subject that is deemed taboo has previously been identified as a feature of euphemistic language change. For example, Allen & Burridge (2006) identified that people use a "general-for-specific substitution", particularly in relation to disease or illness to avoid taboo subjects (e.g. 'mental health' to refer specifically to *mental illness*). They state that this substitution process is a feature of euphemism creation (Allen & Burridge, 2006; see also Grondelaers & Geeraerts, 1998). There is also precedent for the use of general terms to refer to specific subjects in language more generally. For instance, theories of generalised conversational implicature in pragmatics. I discuss implicature and it's role in socially-motivated language change in Section 6.6.

At this point, readers may argue that the increased use in 'mental heath' may not necessarily be indicative that 'mental health' is becoming to mean *mental illness* because a newspaper may report on someone being sectioned under the "mental health act" where clearly the use of 'mental health' refers to *mental illness*. However, further evidence for the interpretation that 'mental health' is being used to refer to a greater number of mental states over the time period is that the increase in the usage of 'mental health' is not linked to the instances of 'mental health' used within a bigger phrase, i.e. 'the mental health act'. The shift in the use of 'mental health' is visible through looking at the relative frequency pmw of 'Mental Health Act' at the start and end of the period for which we have robust data (1985-2013). In 1985 the relative frequency of 'Mental Health Act' is 159.11pmw compared with 23.08pmw in 2013. This again supports the notion that the meaning of *mental health* has broadened, as in the early years of the time period covered, 'mental health' is almost entirely used in reference to the Mental Health Act. Frequency information then is indicative of shifts

in usage but if we are to get a better sense of how the concept of *mental health* has changed, we need to look in more detail at lexical items related to 'mental health'.

6.2. Mental health 'problems', 'conditions' and 'issues'

In the previous section I showed that 'mental health' is increasing in usage. I argued that one explanation for this rise is that 'mental health' is used in the corpus to refer to *mental illness.* To explore this possibility, I will examine the collocates of 'mental health' in more detail, particularly the modifiers of 'mental health'. The reason for doing this is that (as argued in Chapter 3) collocation is a useful analytical method for revealing information about the meaning of a lexical item and how a lexical item interacts with other words. In particular, collocation can be revealing of how take on related meanings by their co-occurence, as suggested by Firth who said of collocation "you shall know a word by the company it keeps" (Firth, 1957: 11). In order to ascertain new usages pertaining to the concept of *mental illness*, e.g. the use of modifiers of 'mental health', I explored the hypothesis that such an analysis would reveal negation in some sense. Table 6.1 shows the top 20 collocates of 'mental health' in the corpus.

Collocate	MI score	Collocate	MI score
1. Problems	8.03	11. People	4.62
2. Act	9.53	12. Sectioned	8.82
3. Issues	8.39	13. Professionals	7.20
4. Charity	7.87	14. Care	5.26
5. Services	7.64	15. Charities	8.02
6. Mind	8.61	16. Service	5.79
7. Foundation	8.82	17. With	4.06
8. Problem	6.33	18. Stigma	7.17
9. Under	6.37	19. Team	5.85
10. Trust	7.08	20. Executive	6.81

TABLE 6.1. TOP 20 COLLOCATES OF 'MENTAL HEALTH' LISTED BY FREQUENCY (CALCULATED USING SKETCH ENGINE, R3-L3, MIN FREQ. = 5, MI CUT-OFF = 3)

Table 6.1 shows that 'mental health' does not collocate with any inherently positive lexical items. The collocates of 'mental health' are either negative ('problems', 'stigma') or neutral ('people', 'act', 'trust'). We can interpret this as demonstrating that 'mental health' can be used to refer to official, formal organisations, e.g. in phrases such as 'mental health charities' and 'mental health trusts', as well as on its own to refer to negative mental states, e.g. 'mental health problems'. Taken with the frequency information for 'mental health' and 'mental illness' presented in Figure 6.1, Table 6.1 indicates further that 'mental health' could be being used to refer to *mental illness* rather than *mental health* generally. The hypothesis that 'mental health' is being used to refer to *mental illness* in the corpus is explored in more detail in Section 6.4 where I explore the concordances and collocates of 'mental illness'.

So far, I have suggested that 'mental heath' conventionally refers to a range of mental states as well as positive mental states. Definitions of 'mental health' support this interpretation. For example the World Health Organisation (WHO) implicitly suggest that *mental health* is a continuum (WHO, 2014), i.e. 'mental health' refers to states of mental wellness as well as states of mental illness, by describing positive mental health as the "positive dimension of mental health". This mental health continuum may be represented in the following way, where the lexical form 'mental

health' refers to the concept of the continuum of mental states as well as the positive dimension of that continuum:

FIGURE 6.2. THE MENTAL HEALTH CONTINUUM AS SUGGESTED BY THE WHO DEFINITION

Further evidence from the corpus that attests my interpretation that 'mental health' is used to refer to the concept of *mental illness,* and therefore that mental health is rarely used the corpus to refer to the positive dimension of the mental health continuum is that 'positive mental health' occurs 51 times in the corpus (0.12 pmw) whereas 'negative mental health' occurs 7 times (0.87 pmw). If 'mental health' were being used to refer to the positive dimension of the continuum then the term would not need to be marked to convey that it referred to *positive mental health, i.e.* 'positive' would be redundant because positive would be denoted by 'mental health'. Battistella (1990) defines semantic markedness as "a relation between a very specific linguistic sign (the marked term) and a sign that is unspecified for the grammatical or conceptual feature in question" (Battistella, 1990: 2). The fact that the 'mental health' needs to be marked to convey positive mental health then, provides further linguistic evidence that the concept of *mental health* does not encode positive mental states in the corpus.. The need to use the marked form is interesting because the definition of *mental health* given by WHO (2014) above does not preclude its use to refer to positive mental states. This suggests that 'mental health' is being used to refer to a concept (i.e. *mental illness)* that is semantically more specific that its dictionary definition suggests it can be. Battistella (1990) writes of contextual markedness that "markedness values are also

contextualised within a language. Values are not fixed, but rather are relative: cultural and linguistic structure acts as a context within which categories are evaluated" (Battistella, 1990: 24). What Battistella (1990) is referring to in this quotation is how markedness can reveal the sense of a lexical item within a particular culture or discourse. The example he gives to exemplify context dependent markedness is the noun 'nurse' which can refer to both male and female nurses, but which is conventionally marked when it is used to refer to a man, i.e. 'male nurse'[66]. This marked form reveals that within the context of nursing (at least at the time Battistella was writing), 'nurse' usually refers to a female nurse. Markedness then, can be revealing of what is considered the norm within a particular context. To return to the mental health example, the modification, or markedness of "positive mental health" suggests that within the context of mental health reportage, 'mental health' is not used to refer to positive states of mental health.

To explore any other terms that may be used to modify 'mental health' to convey negative states, I searched for 'problem' using the Sketch Thesaurus feature on Sketch Engine (Kilgariff et al., 2014). The Sketch Thesaurus function uses the corpus to generate usage-based synonyms, i.e. words that occur in the same, or very similar linguistic contexts which may not be conventional synonyms but which are being used within the corpus in similar contexts. The reason for exploring 'problem' specifically was that it was the most frequent collocate of 'mental health' (as shown in Table 6.1). Further, the idea behind exploring 'problem' in more detail using the Sketch Thesaurus function was that it would reveal any other nouns that post-modified 'mental health' that were similar to 'problem' that could then inform further searches for terms conveying states of negative mental health, e.g. 'issues' or 'disorders'. This search yielded the additional terms 'condition', 'symptom' and 'disease'. 'Mental health disease' and 'mental health symptom' were discounted on the basis that only a few instances of each term occurred in the corpus. Arguably,

[66] My intuition is that this usage has become less frequent since 1990 when Battistella was writing.

however, the fact that these terms feature at all indicates that the meaning of 'mental health' is changing, specifically that *mental health* is not conceptualised as a continuum, but rather as a state. Evidence that the meaning of 'mental health' is changing to refer to a specific mental state (i.e. illness) and not the continuum of mental states is that there are 18 instances in the corpus of 'mental health symptom' which, if we consider mental health to refer to the continuum of mental states is semantically incongruous as one cannot have a symptom of a continuum. A 'symptom' of mental health is semantically under-specific unless 'mental health' in the 18 instances is being used to refer to a state of mental illness. To explore the patterning of 'mental health problem*', 'mental health condition' and 'mental health issue' across the time period, relative frequencies were plotted for each term. Figure 6.2 shows the relative frequency of 'mental health problem', 'mental health condition' and 'mental health issue' over the time period covered in the MI 1984-2014 corpus. In order to show the relative frequency of 'mental health issue' and 'mental health condition' clearly, Figures 6.3 and 6.4 show the relative frequency of 'mental health issue' and 'mental health condition' respectively. To avoid cases where 'mental health issue' was used to relate to something other than a mental state, only 'mental health issue' was searched for, and not 'mental health issues*' which may have returned hits pertaining to usages such as "the Government's record on mental health issues is the total antithesis of its alleged intention", where mental health issue refers to schemes surrounding mental health and not a mental state.

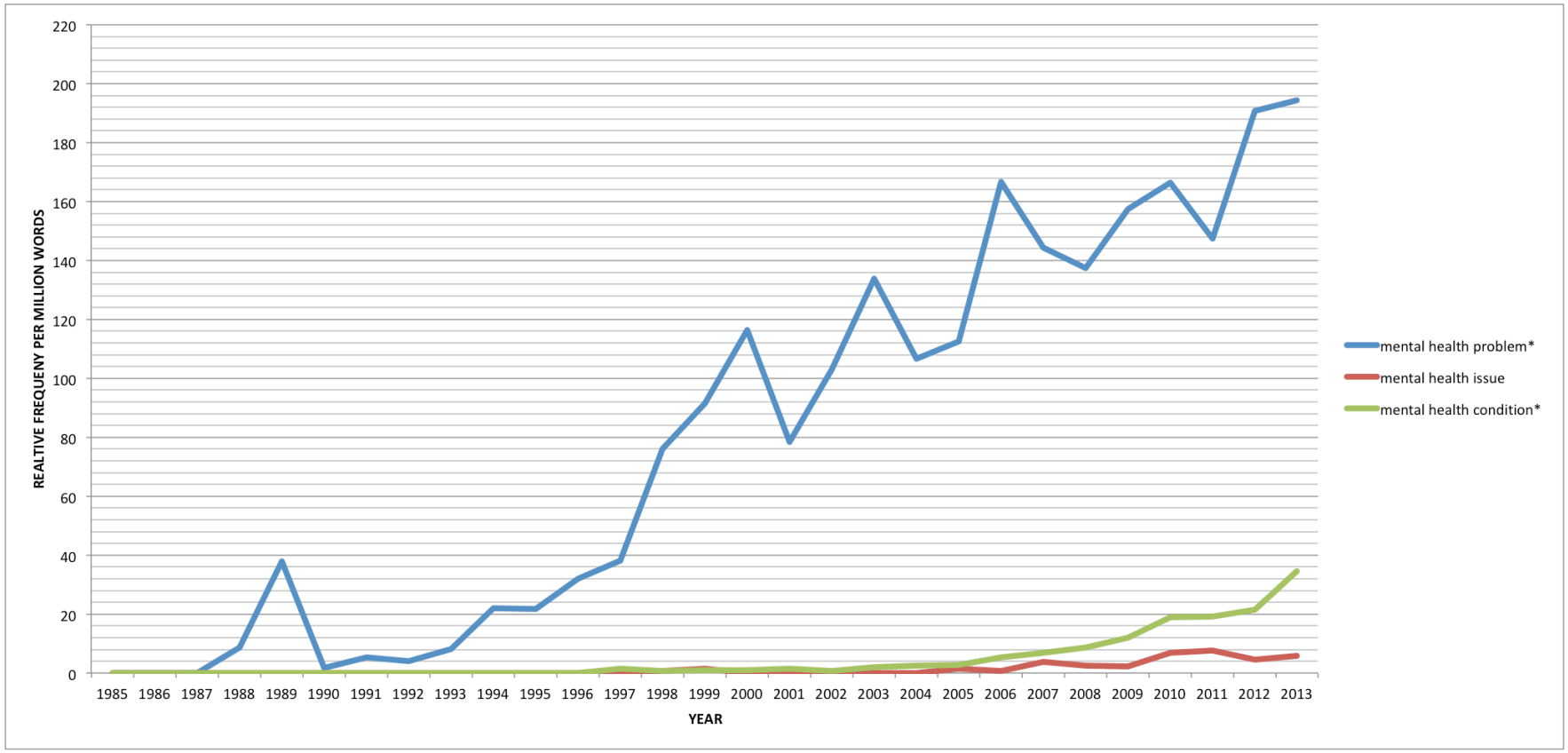


FIGURE 6.3. FREQUENCY OF 'MENTAL HEALTH PROBLEM/ISSUE/CONDITION' OVER THE TIME PERIOD.

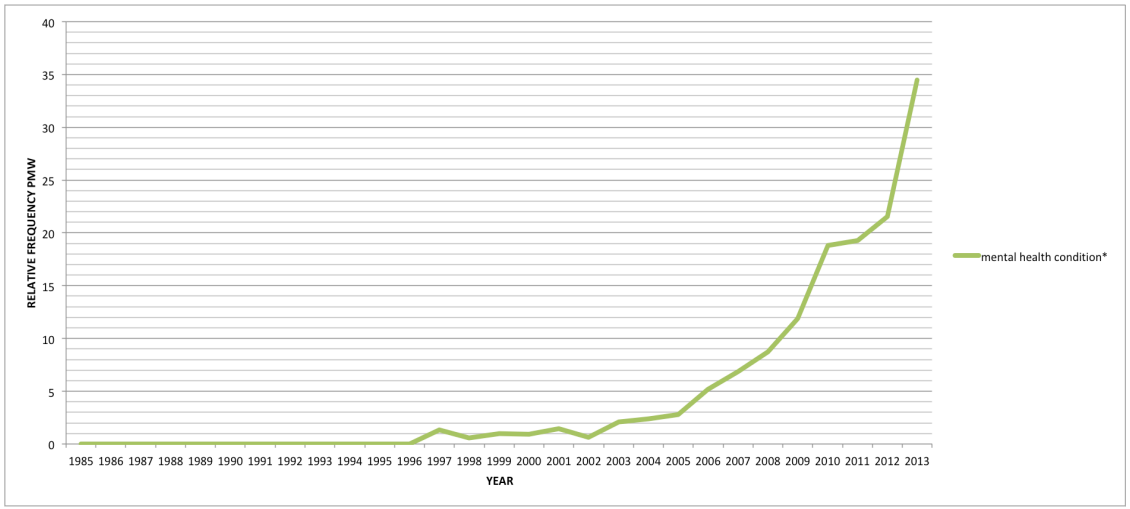


FIGURE 6.4. FREQUENCY OF 'MENTAL HEALTH CONDITION OVER THE TIME PERIOD

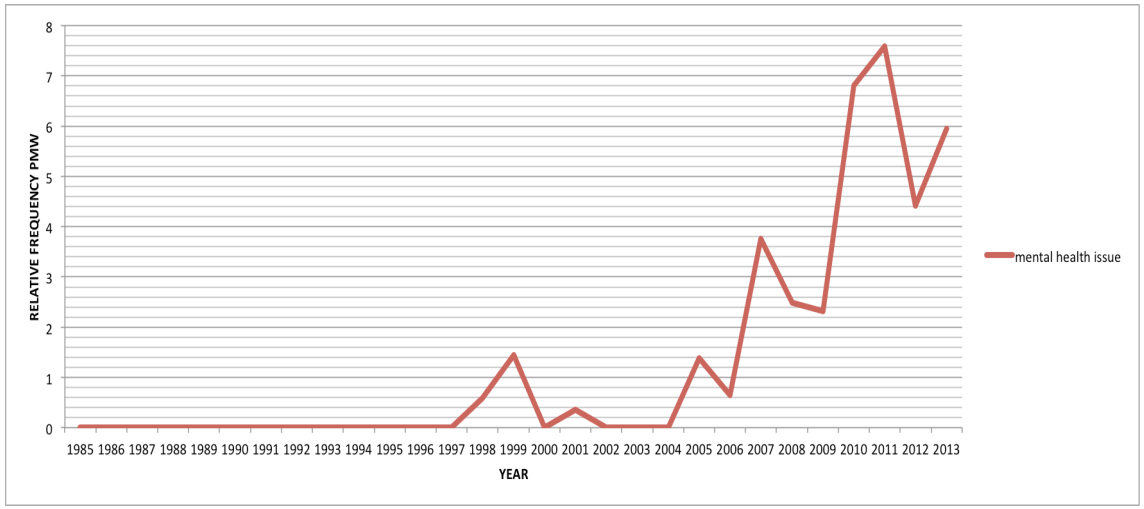


FIGURE 6.5. FREQUENCY OF 'MENTAL HEALTH ISSUE' OVER THE TIME PERIOD

Figures 6.3, 6.4 and 6.5 show that we have developed, and are developing, increasingly common terms to convey the concept of *negative mental health*, i.e. the state of being mentally unwell. A closer inspection of the concordances of these terms further indicates that 'mental health' and 'mental illness' are used as near synonym (see Table 6.2). This can be seen through anaphoric and cataphoric reference whereby 'mental health issue' is used to refer to a mental illness, e.g. OCD, or the lexical item 'mental illness'. The first instance occurred in 2001 with the second two instances occurring in 2005.

According to Watters, traditional ways of treating **mental illness** are not always best for asylum seekers. "It takes a lot of time to work through a **mental health issue** and you need a degree of stability to do it
obsessive compulsive disorder is the most common **mental health issue** after depression
One in four people will suffer from a **mental health issue** such as **depression** at some point in their life

TABLE 6.2. CONCORDANCE LINES FOR 'MENTAL HEALTH ISSUE'

These concordances and Table 6.2 give further support to the notion that 'mental health' is now used to refer to states of *mental illness*. Moreover the fluctuating frequencies of 'mental health issue/problem/condition' and the fact that they start to occur around the same period (c.2000-2004) suggests that we can expect the frequency of these terms to continue to rise after 2013 as they are becoming established terms within the semantic domain of MENTAL HEALTH AND ILLNESS. Further evidence for this claim can be found by identifying the total number of instances of 'mental health' in the corpus and calculating what percentage of those total instances refer to cases where 'mental health' is used with 'condition', 'problem' or 'issue'. Figure 6.6 shows the overall trend for 'mental health' to be used within the phrases 'mental health problem', 'mental health illness' and 'mental health condition'.

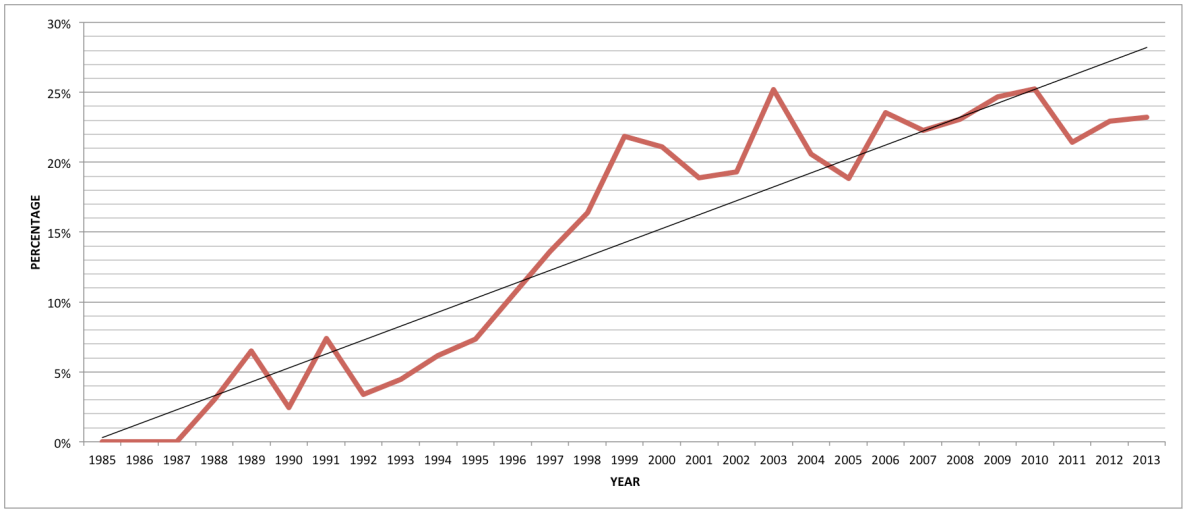


FIGURE 6.6. PERCENTAGE OF INSTANCES OF 'MENTAL HEALTH' THAT OCCUR IN THE PHRASES 'MENTAL HEALTH PROBLEM', 'MENTAL HEALTH ILLNESS' AND 'MENTAL HEALTH CONDITION'

Figure 6.6 shows that the usages of 'mental health' to indicate *negative mental health* has increased from zero in 1985 to 25% in 2010. This result provides evidence that this shift is happening, and the terms within the semantic domain of MENTAL HEALTH AND ILLNESS are shifting towards the negative end of the continuum.

6.3. The rise of 'wellbeing'

As I stated in the introduction to this chapter, the analysis of new lexis added to (or emerging in) a semantic domain or discourse provide insight into what concepts the existing terms in that semantic domain are taken to refer to (because the addition of new words may indicate that existing words in that semantic domain cannot convey some meaning) . So far, I have argued that in the corpus there is a strong tendency for 'mental health' to refer to the concept of *mental illness*. I have argued that this broadening of meaning is a result of euphemistic language use. Furthermore, I showed that 'mental health' is marked in the corpus to refer to the concept of *mental wellness* (e.g. "positive mental health"). If we accept that the meaning of 'mental health' has broadened and shifted to refer to the concept of *mental illness* then we might predict that another term is emerging in the semantic domain of MENTAL

HEALTH AND ILLNESS that only refers to the concept of *mental wellness* (because 'mental health' has become associated with the concept of mental illness and is therefore less frequently used to refer to mental wellness). I explore this process of language change in more detail using linguistic evidence that demonstrates this change (Traugott & Dasher, 2002) in Section 6.6, but for now I will use this prediction to explore what new terms are emerging to refer to *mental wellness.*

In order to ascertain whether a term has emerged to convey the concept of *mental wellness*, I used the Word Sketch function on Sketch Engine (Kilgariff et al., 2014) which shows statistically significant collocates of a query term (grouped by frequency of the collocation, using the logDice statistical calculation). I explored the syntactic frame '[mental health and]' to see which lexical items followed. I did this on the basis that the additive property of 'and' grammatically connects two NPs, creating one subject, object or complement comprising a complex NP or NP embedded in a PP. Doing this then provided a selection of refined terms for concordance analysis as it allowed for closer, more specific inspection of the surrounding text of '[mental health and…]'. Focussing in on the data means that is it easier to conduct more specific, qualitative analysis, as it is possible to see whether the semantic content of the surrounding context of '[mental health and…]' suggests that the two lexical items were being equated discursively. Through this analysis, I established that the two most common items following '[mental health and…]' was 'wellbeing' (98 instances, 1.67 pmw) and 'well-being' (45 instances, 0.77 pmw). An example of the concordance of 'mental health and wellbeing' (top six randomly generated examples in the corpus) is shown in Table 6.3.

Aspects of life vital for	**mental health and wellbeing**	- sleep, diet, exercise, recreation and relationships
As a nation, we have much to gain from an investment in	**mental health and wellbeing:**	confidence, resilience and the improved ability of our people translate into greater opportunities
The government's recent inquiry into	**mental health and wellbeing**	in later life estimates 2.4 million older people in Britain suffer from depression
'This report highlights the fact that the	**mental health and wellbeing**	of individuals, not only from the farming community but also from other rural businesses and those working to tackle the outbreak on the frontline, was affected and these effects may go on for some time
'Projects such as this fit very much with the aims of the Executive's national programme for improving	**mental health and wellbeing**	-
Weeks after John died, Isabel and her son Hugh planted the seeds of Theatre Nemo a charity-based theatre group to promote good	**mental health and wellbeing**	while aiming to challenge stigma

TABLE 6.3. CONCORDANCE OF FOR 'MENTAL HEALTH AND WELLBEING'.

Table 6.3 shows that at some level, 'mental health' and 'wellbeing' are being semantically linked because in all the examples the surrounding context describes both 'mental health' and 'wellbeing', i.e. "sleep, diet, and exercise" are vital for both 'mental health and wellbeing' and "investment in the areas of mental health and wellbeing" result in the same outcome: "confidence, resilience". 'Mental health' and

'wellbeing', then, are discursively constructed as being closely related. Furthermore, in concordance line 3 in Table 6.3, an "inquiry into mental health and wellbeing" is reported as being linked to statistics regarding the incidence of depression in later life. This suggests again that 'mental health' and 'wellbeing' are being linked, namely, that the concepts of *mental wellness* and *wellbeing* are being linked.

'Wellbeing' is a potential candidate to take on some of the sense of *mental wellness* in the absence of a term to denote *mental wellness* only (because 'mental health' now increasingly refers to the concept of *mental illness*). The reasons for this hypothesis include the fact that 'wellbeing' relates only to a state of mental wellness (as opposed to illness) and is associated primarily with the maintenance of that state (i.e. 'wellbeing' only refers to the positive end of the mental health continuum). Dodge et al. stated in 2012 that 'wellbeing' was still largely undefined (2012: 222) which attests to its relative newness as a concept within the time period being analysed in this thesis. The newness of *wellbeing* is also attested by its frequency over the time period, which shows that despite low raw frequencies, the term is rising significantly from around half way through the time period. Figure 6.7 shows the frequency of 'wellbeing' over the time period.

FIGURE 6.7. FREQUENCY OF 'WELLBEING'[67] OVER THE TIME PERIOD

The discursive meaning of 'wellbeing' can be investigated using the Thesaurus function on Sketch Engine (Kilgariff et al., 2014), which uses the collocates of a query word to generate a list of synonyms based on usage. Figure 6.8 is the Sketch Thesaurus visualisation for 'wellbeing' and 'mental health'. The visualisation shows that 'wellbeing' is used to describe positive emotional states, e.g. 'happiness', but also to discuss methods for maintaining a state of positive mental health. It also shows that 'health' (used here to refer to 'mental health' because Sketch Engine does not allow for multiword searches) appears to collocate with words to do with illness, which supports the idea detailed above that 'mental health' is now used to refer to *mental illness*.

[67] Figure 6.9 includes instances of both 'well-being' and 'wellbeing'.

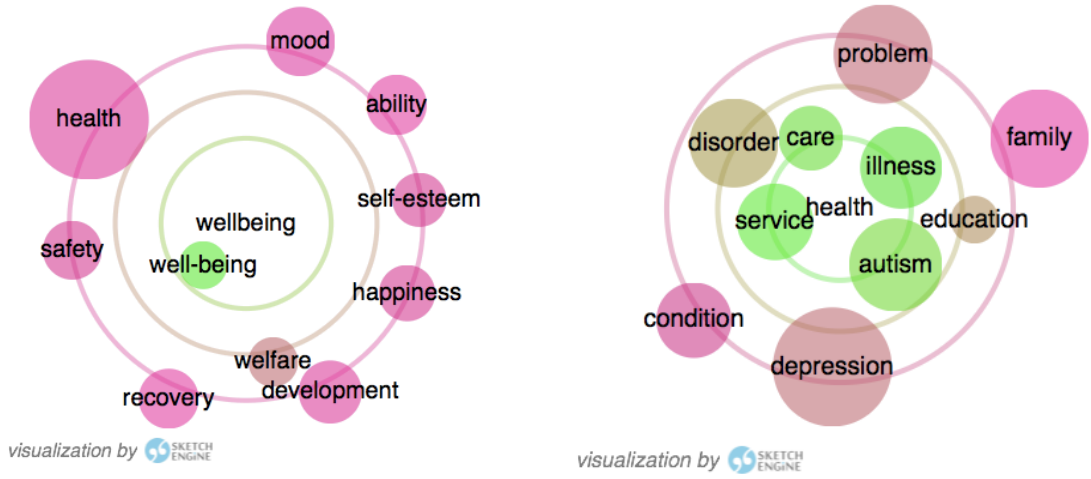


FIGURE 6.8. SKETCH THESAURUS VISUALISATIONS FOR 'WELLBEING' AND 'HEALTH'

From Figure 6.8 it is possible to see how the meaning of 'wellbeing' is associated with lexical items that one may associate with *mental wellness*, e.g. 'happiness' and 'self-esteem'. In contrast, 'health' is associated with more negative lexis such as 'problem', 'disorder', 'illness' and 'condition'. This supports the idea that 'wellbeing' is concerned with the concept of *mental wellness* and 'mental health' is concerned with the concept of *mental illness*. Another indication that 'wellbeing' is becoming established as a term within mental health discourse is that there is evidence that it is becoming lexicalised (where a new word is added to the lexicon) via a process of compound fusion, where morphological boundaries are erased resulting in "unified lexemes over time" (Brinton & Traugott, 2005: 44). Brinton & Traugott state that such lexicalisation is commonly the result of institutionalisation which refers to "the spread of a usage to a community and its establishment as the norm" (Brinton & Traugott , 2005: 45). Information about the spelling conventions of 'wellbeing' provide evidence for the possibility that 'wellbeing' is becoming lexicalised. In 1987, 75% of instances of wellbeing were spelled 'well-being', however in 2013 'wellbeing' became the conventionalised spelling with 74% of all instances spelled this way. This convention in spelling is a strong indication that wellbeing is being lexicalised, i.e. added to the lexicon. Figure 6.9 shows this process of lexicalisation using percentages of the total

number of 'wellbeing' and 'well-being'. Figure 6.10 shows the raw frequencies of the two terms.

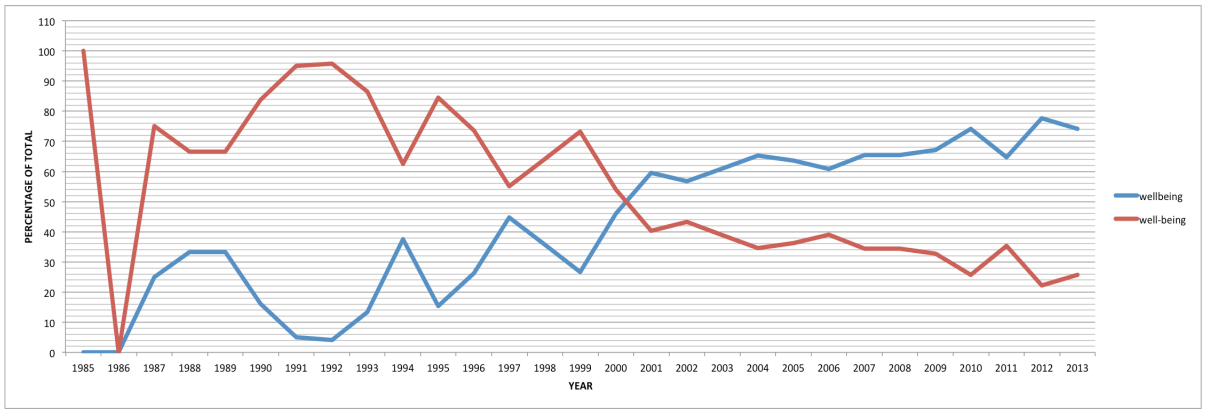


FIGURE 6.9. PERCENTAGE DIFFERENCES BETWEEN 'WELLBEING' AND 'WELL-BEING'

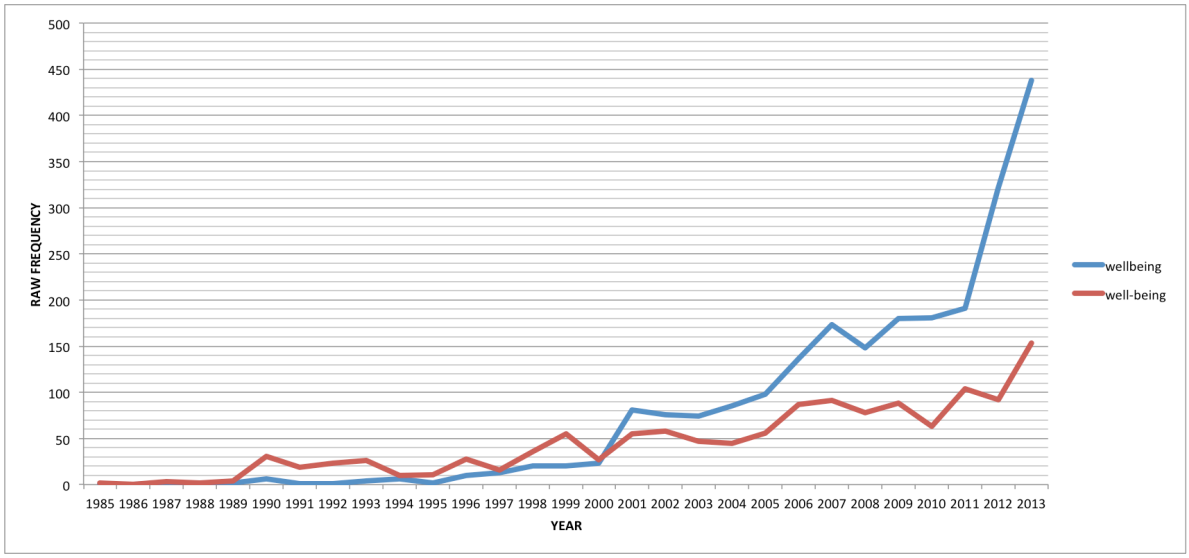


FIGURE 6.10. FREQUENCY OF 'WELLBEING' VS. 'WELL-BEING'

Despite fairly low frequencies of these lexical items, Figures 6.9 and 6.10 show that at the year 2000, 'wellbeing' became the dominant spelling. Since 2000, the use of 'wellbeing' has been increasing in line with other terms to do with *mental illness* shown in Figure 6.1. Given that the data contained in the MI 1984-2014 corpus is newspaper data (and newspapers have style guides), it could be the case that 'wellbeing' is being spelled using both variants due to style guides for individual publications; however, if that were the case we could expect to see 'well-being' drop out of use as style guides

adopt 'wellbeing' only, but it does not. We can hypothesise, then, that in future 'wellbeing' will become the only form used. Taken together, the lexicalisation in progress of 'wellbeing' and the increase in its usage support the argument that new terms are becoming established within the discourse of mental health and illness, i.e. a set of semantically related and static terms to refer to the continuum of mental health.

6.4. Assessing similarity and difference through collocation

In Sections 6.1 through 6.3, I used frequency analysis, concordance analysis, and collocation analysis to explore similarities and differences in the terms 'mental health', 'mental illness' and 'wellbeing'. In this section I focus more specifically on looking at shared and distinct collocates of terms as a means of identifying similarities and differences in the labels that have been identified in Sections 6.2-6.3. To do this, I use the Word Sketch Difference tool on Sketch Engine (Kilgariff et al., 2014), which shows the shared and distinct collocates for two query items. The Word Sketch Difference function is useful for looking at nuanced differences in items that share a semantic field, within which similarities between terms may be harder to evidence using frequency and concordance analysis. Figure 6.11 shows the Word Sketch Difference for 'illness' and 'health' in the corpus.

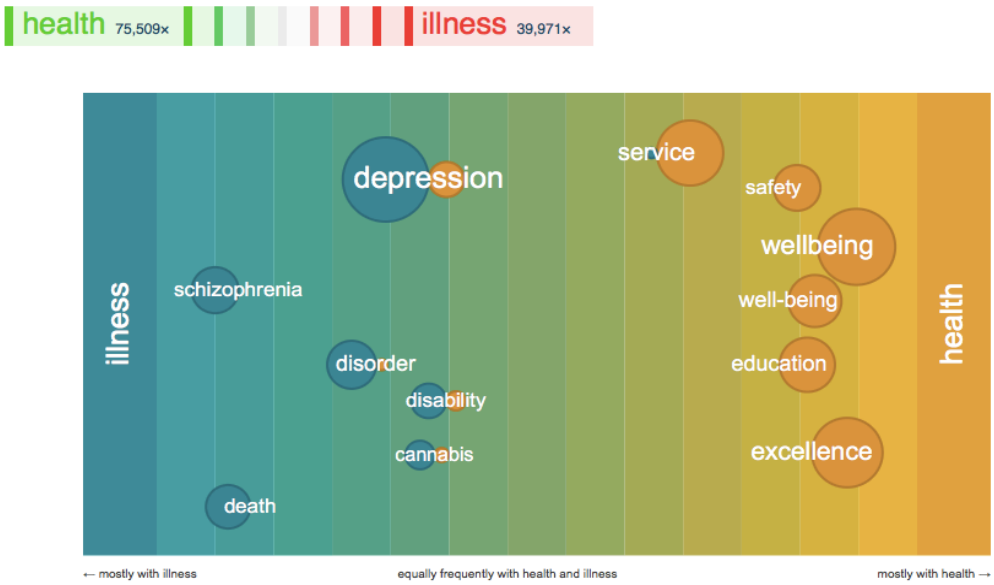


FIGURE 6.11. WORD SKETCH DIFFERENCE FOR 'ILLNESS' AND 'HEALTH'

Figure 6.11 shows, in line with the argument made so far, that the meaning of *wellbeing* is closely related to *health* (evidenced by the fact that 'wellbeing' and 'well-being' collocate with 'health'. So far, I have demonstrated this through qualitative concordance analysis and relative frequencies. The collocation analysis provided by the Word Sketch Difference tool is another indication that the two concepts potentially share meaning. Figure 6.12 shows the Word Sketch Difference for 'problem' and 'illness'. The reason for this analysis is that, prior to this point, I have argued that 'mental health problem' is a candidate term to convey the concept of *mental illness*. Knowing this, one might expect to see more negatively-valenced lexical items at the 'problem' end than at the 'illness' end. The reason that we might expect to see more negatively-valenced lexical items at the problem end is that the number of articles from later years is greater than those in the early years in the corpus, and as a result, the shift from 'mental illness' to 'mental health problem' would be more pronounced when looking at the corpus as a whole to reflect this change in the meaning of 'mental health problem' and 'mental illness'.

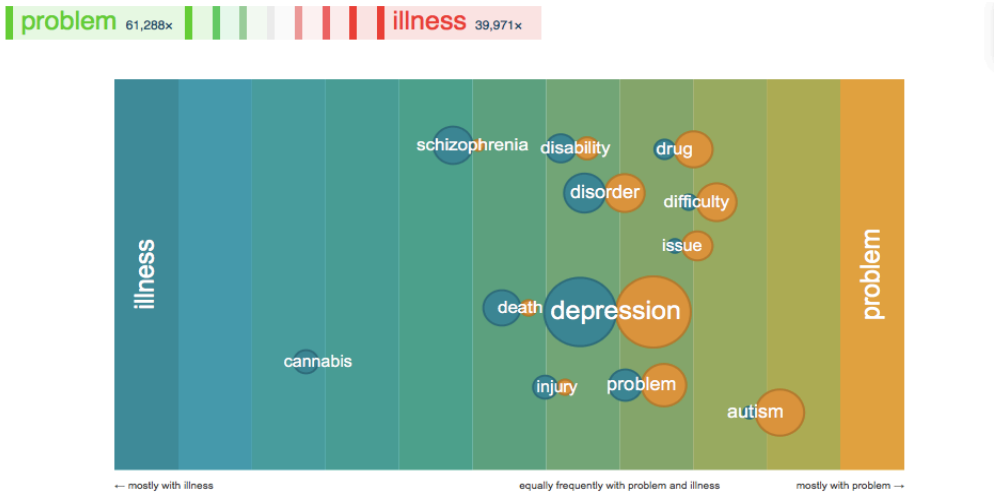


FIGURE 6.12. WORD SKETCH DIFFERENCE FOR 'ILLNESS' AND 'PROBLEM'

Looking at Figure 6.12, it appears that the hypothesis regarding 'mental health problem' bears out. As can be seen, most of the terms included occur equally frequently with both terms. Using collocation as a benchmark for usage-based synonymy, then, we can be fairly confident that these two items are have some shared sense. This indication of usage-based synonymy is also echoed in Figure 6.13, which shows the Word Sketch difference for 'health' and 'wellbeing'.

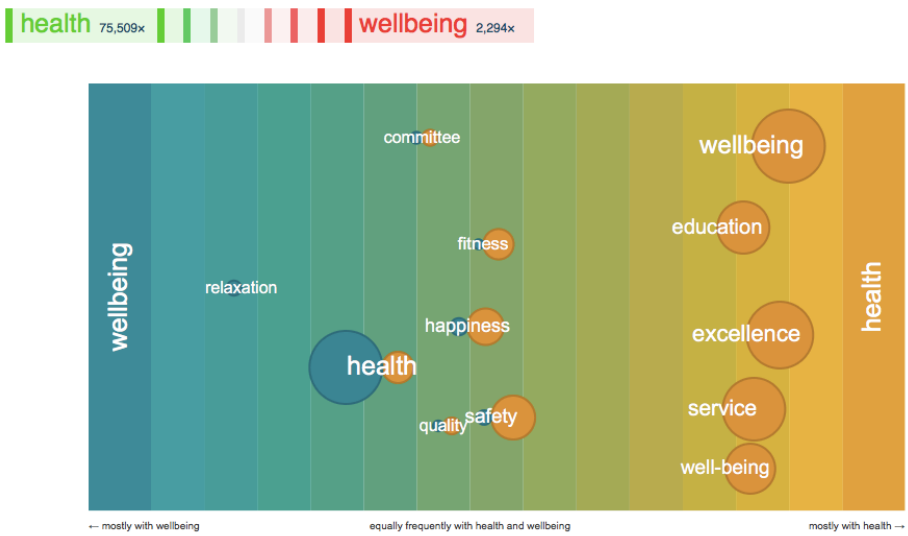


FIGURE 6.13. WORD SKETCH DIFFERENCE FOR 'WELLBEING' AND 'HEALTH'

Figure 6.13 shows that 'wellbeing' and 'health' share a large number of collocates that occur equally frequently, which again could suggest that, when taken together with qualitative analysis, these terms are usage-based synonyms within this discourse context, or at least closely related within a shared semantic field. 'Wellbeing' then, appears to be a good candidate term to convey *mental wellness* in light of the fact that 'mental health' is being used to refer to *mental illness* via nominal modification such as 'problems' and 'issues'. Moreover, 'wellbeing' has been added to the discourse of mental health and illness within a short and recent period of time, which indicates that 'wellbeing' has been taken up in this context by UK journalists, writing for people within UK society. This in turn suggests that *wellbeing* is a recognisable concept for people within UK society.

If we step back for a moment from the linguistic analysis conducted here and consider the idea of the stigma around mental illness that is often mentioned in the literature reported in Chapter 2, arguably, 'wellbeing' is a good candidate term to denote *mental wellness* as no link exists in the MI 1984-2014 corpus between 'wellbeing' and 'stigma'. For example, in a search of collocates of 'stigma', both 'health' and 'illness' appear as collocates, whereas no such relationship occurs between 'wellbeing' and 'stigma'. In fact, there were no instances in the corpus where 'stigma' occurred with 'wellbeing'. Tables 6.3 and 6.4 show example concordances (top three randomly generated concordance lines) in the corpus for 'stigma' and 'mental health'/'mental illness'.

tackle the	**stigma**	of **mental health**
because of the	**stigma**	attached to **mental health**
break the	**stigma**	around **mental health**

TABLE 6.3. CONCORDANCE FOR 'STIGMA' + 'MENTAL HEALTH'

break down	**stigma**	attached to **mental illness**
hopes to address the	**stigma**	surrounding **mental illness**
aiming to end the	**stigma**	attached to **mental illness**

TABLE 6.4. CONCORDANCE FOR 'STIGMA' + 'MENTAL ILLNESS'

Furthermore, if we look at the Word Sketch collocates returned for the syntactic frame '[wellbeing and…] (as I did previously in Section 6.3 to provide some evidence for a link between 'mental health' and 'wellbeing'), we see that 'wellbeing' is routinely associated with terms that denote a positive state of mind and maintaining a positive state of mind.

Wellbeing and…	
1. health	11. euphoria
2. happiness	12. confidence
3. relaxation	13. comfort
4. safety	14. function
5. committee	15. future
6. fitness	16. mood
7. quality	17. relationship
8. self-esteem	18. session
9. satisfaction	19. emotion
10. morale	20. development

TABLE 6.5. WORD SKETCH COLLOCATES OF "WELLBEING AND…"

Moreover, the fact that 'wellbeing' occurs in phrases such as "wellbeing committee" provides further evidence that this term is being adopted in an official capacity because it is being recognised by groups of people in an institutional setting.

In the next section I explore the idea that the terms 'mental illness', 'mental health' and 'wellbeing' have positive or negative associations, using the concepts of semantic preference and semantic prosody (see Chapter 3, Section 3.3.8 for a description of semantic preference and semantic prosody).

6.5. The collocational context of 'mental health', 'mental illness' and 'wellbeing'

In Section 6.4, the terms 'mental health', 'mental illness' and 'wellbeing' were explored using the Word Sketch function. This analysis offered insight into whether terms were more positive or negative than a comparison term. In this section, I conduct a more specific analysis of the positive or negative associations of each term separately. Table 6.3 shows the top 10 collocates of each term.

Mental illness	Mental health	Wellbeing
stigma, severe, serious, rethink, suffering, history, suffer, people, form, attached	problems, act, issues, charity, services, mind, foundation, problem, under, trust	emotional, overall, physical, mental, health, improve, suffolk, sense, psychological, promote

TABLE 6.6. TOP 10 COLLOCATES OF 'MENTAL ILLNESS', 'MENTAL HEALTH' AND 'WELLBEING' (CALCULATED USING SKETCH ENGINE).

Table 6.6 reinforces my argument that 'mental health' is now being used to refer to *mental illness* (as indicated by 'problems') it also reinforces the assessment made in Section 6.1 that 'mental illness' can also refer to official organisations, such as "Trusts" and "Foundations". Similarly, Table 6.6 supports the point made in the previous section that 'wellbeing' is also becoming established through official campaigns, e.g. *"Suffolk Health and Wellbeing month", "had been awarded the Suffolk wellbeing service contract", "a clinical psychologist with the Suffolk wellbeing service"*

Table 6.6 also suggests that *mental illness* is viewed as the most negative concept of the three. Evidence for this is that the collocates of 'mental illness' have a negative semantic prosody whereby lexical items in particular units of meaning take on negative associations due to the negative meaning of the words they collocate with, e.g. one does not "suffer" from something positive, "stigma" does not surround something good, and good things don't tend to be described as "severe" or "serious". This last point is attested by a collocation analysis of 'severe' in the British National Corpus (BNC) in which the top 10 collocates (ranked by frequency, L4-R4) of 'severe'

include 'problems', 'damage', 'difficulties', 'suffered' and 'patients', and the top 10 collocates of 'serious' include 'problems', 'injury', 'threat, 'offence and 'damage. The table also suggests that compared with 'mental illness', 'mental health' is viewed as a more neutral term (although this is shifting towards being a more negative term indicated by 'problem' and 'problems') and that 'wellbeing' is a more positive term. Evidence for the view that 'wellbeing' is more positive than 'mental illness' and 'mental health' is that the collocates of 'wellbeing' have a more positive semantic meaning. For example, in the BNC the top 10 collocates (ranked by frequency, L4-R4) of 'promote' include 'development, 'health, 'interests', 'growth' and 'awareness'. The fact that these words are positively valanced suggests that the action of 'promoting' is seen positively. For instance, 'actively' usually premodifies verbs that convey positive aims. As an example, in the BNC 'actively' premodifies 'engaging', 'pursuing', 'participating' and 'growing'. Another example of the positive semantic prosody associated with 'wellbeing' is the collocate 'sense' which, under closer concordance analysis, forms part of a bigger unit of meaning – "a sense of". The collocates of this unit of meaning in the BNC are 'humour', 'belonging' and 'identity', which again convey positive emotions, particularly around the notion of affirmation of oneself. Another interesting insight into 'wellbeing' and how it fits into the continuum of mental health discourse is provided by 'improve' (see Table 6.6). The collocates of this lexical item in the BNC are 'efficiency', 'performance', 'quality', 'relations' and 'standards', which appear superficially to fit neither a negative nor a positive semantic prosody. However, if we attempt to categorise these words by their semantic preference we can see that several of these terms invoke the idea of a continuum, i.e. all these words describe a series of states. Moreover, from a semantic perspective, often the word used to indicate a continuum also performs the function of denoting the positive end of that continuum; e.g. 'quality', which both refers to the continuum of quality and describes the state of something being good quality. This phenomenon was demonstrated in Figure 6.2 ('the mental health continuum') which was based on the WHO definition of mental health, where 'mental health' was a term to describe

both the continuum and, in the absence of another term, the positive end of the scale. The reason such words are interesting is that their occurrence with 'wellbeing' suggests that 'wellbeing' triggers a continuum and yet evidence in my analysis so far suggests that 'wellbeing' is viewed as a wholly positive thing. This raises interesting questions about the future of 'wellbeing' as a positive term, as it may be the case that as it becomes adopted into common parlance, new ways of negating its meaning of *a positive mental state* may be created, as has been the case for *mental health*. There is however, a vital difference between the labels 'mental health' and 'wellbeing'. This difference is that 'mental health' is associated with mental states by virtue of its lexical construction, i.e. the word 'mental' is a constituent of the lexical item. For this reason, 'wellbeing' may well be a good candidate for anti-stigma initiatives in light of 'mental health' shifting towards negative associations. This is because wellbeing does not have a marked form to denote negation, and nor does it have a conventionalised antonym, e.g. 'unwellbeing' and 'non-wellbeing' do not exist in the English Language. Consequently, another term would have to be adopted that is lexically (as opposed to semantically) distinct from 'wellbeing', which would mean that 'wellbeing' would not be contained in the term to denote *opposite wellbeing*. This argument explains to a certain extent why 'mental health' has shifted to the negative end of the continuum of mental health discourse because health has a conventionalised negative form – 'illness'.

6.6. The contemporary view of mental health: pragmatic accounts of language change

So far I have argued that the data suggests 'mental health' is increasingly used as a euphemistic term to refer to the concept of *mental illness*. I have provided evidence for this argument by showing that 'mental health' is increasingly being modified to convey this meaning, e.g. *mental illness*. I have also showed that 'wellbeing' is used in the corpus to refer to the positive dimension of *mental health.*. What I argue in this

section is that this diachronic change, whereby 'mental health' can now refer to *mental illness*, is entirely consistent with patterns of language change around euphemistic terms, whereby semantic change is instigated by "the strategic use of language for communicative purposes" (Traugott & Dasher, 2002: 58). Furthermore, euphemistic language use has been identified as an instigator of language change due to the fact that taboo and euphemisms "provide an emotion trigger for word addition, word loss, phonological distortion and semantic shift" (Burridge, 2012: 65). Furthermore, euphemistic usages result in semantic language change because

> words may come to be perceived as unpleasant […] because they are linked to some […] culturally sensitive material or behaviour. This may trigger tabooing and subsequent loss of the original term and/or euphemistic extension of another item
>
> (Urban 2015: 375)

A pragmatic account of lexical change can be used to explain the euphemistic use of 'mental illness'. Such a shift in usage is motivated by social-cultural impulses (Traugott, 2010: 551; Urban, 2015), specifically our desire to avoid taboo subjects by the use of euphemisms. The taboo nature of *mental illness* is attested by the fact that the syntactic frame '[taboo is…]'[68] returns instances of 'mental illness is a taboo'. Taboo and euphemism were identified by Bréal (1964) as motivations for language change. In essence, the meaning of the word being used euphemistically is broadened such that the original taboo referent is gradually lost. Traugott and Dasher (2002) illustrate Bréal's point using the example of the lexical item *toilet*, which broadened in meaning from referring to the cloth used to wrap one's head, to the activities associated with grooming generally, and finally to "the fixture for disposing of bodily excretions, and the room containing it" (Traugott & Dasher, 2002: 59). The usage of *toilet* to refer to activities associated with grooming declined because of its status as a

[68] This syntactic frame captures both 'X is a taboo' and 'taboo is X'

euphemism for things to do with excretion. Then, because *toilet* became more associated with excretion than grooming, it became taboo again, so *toilet* was replaced with "terms such as *restroom*, or *bathroom* (even when no bath is expected or known to be present) (Traugott & Dasher, 2002: 59). Figure 6.14 shows the process of pragmatic-led semantic change for 'toilet'.

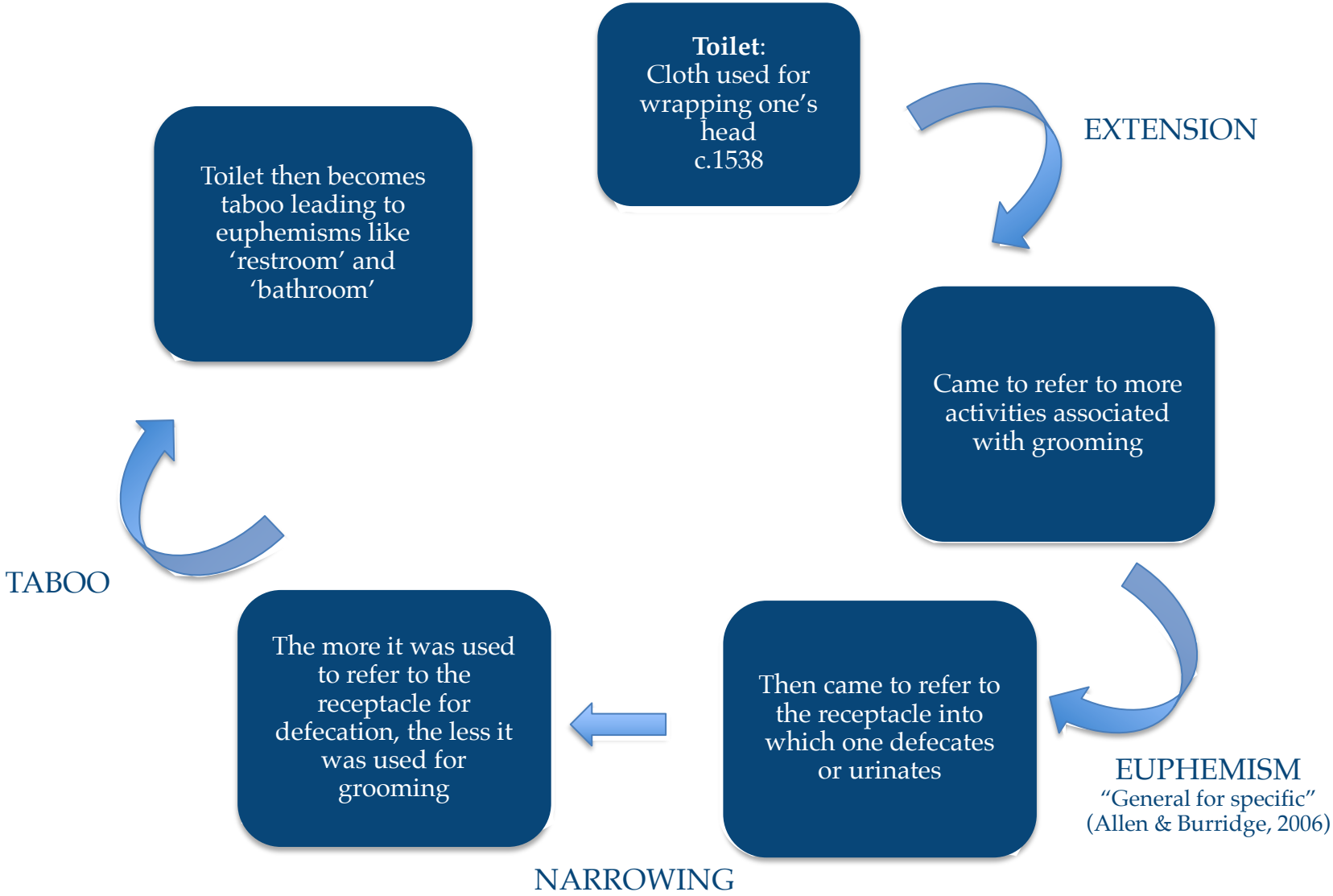


FIGURE 6.14. PROCESS OF SOCIALLY-MOTIVATED LANGUAGE CHANGE FOR 'TOILET' (TAKEN FROM TRAUGOTT & DASHER, 2002: 59)

The case of the changing meanings of *toilet* described by Traugott & Dasher (2002) provides a clear example of how the avoidance of taboo subjects via euphemism is managed in language. However, the corpus used in my thesis does not cover a large enough time span for me to be able to document a complete cycle of euphemistic language change. The first recorded use of *toilet* in the first sense that Traugott and Dasher (2002) describe was in 1538. Such a change, then, happens over a considerable period of time, and our understanding of the changing meaning of lexical items is always at least partially retrospective. However, I argue that the fact that we can observe change within the 30-year period documented in the corpus demonstrates the significance of the changes I have identified. Moreover, such findings show how powerful (corpus) linguistic tools are for tracking emergent semantic change in electronic data. I cannot claim to have witnessed a full cycle of change in the data available, where 'mental health' refers to something far removed from mental states. What I can claim, however, is that, societally, 'mental illness' is viewed as taboo to a certain extent (evidenced by its co-occurrence with words like 'taboo' and 'stigma)', and that 'mental health', which did not have negative associations during the early years of the time period, was used as a euphemistic term to refer to *mental illness*. 'Mental health' was then modified to refer more obviously to *mental illness* via the addition of 'mental health problem'/'issue'. Considered in relation to the 'toilet' example of semantic change over centuries , this change within thirty years is significant and provides evidence for the fact that the contemporary discourse of mental health and illness is still emerging. Figure 6.15 shows this emergent language change using the processes of language change identified by Traugott & Dasher (2002). To reflect that the diagram is based in part on prediction, the two leftmost boxes in the diagram are marked in gradient colour to show that this possible language change is still ongoing.

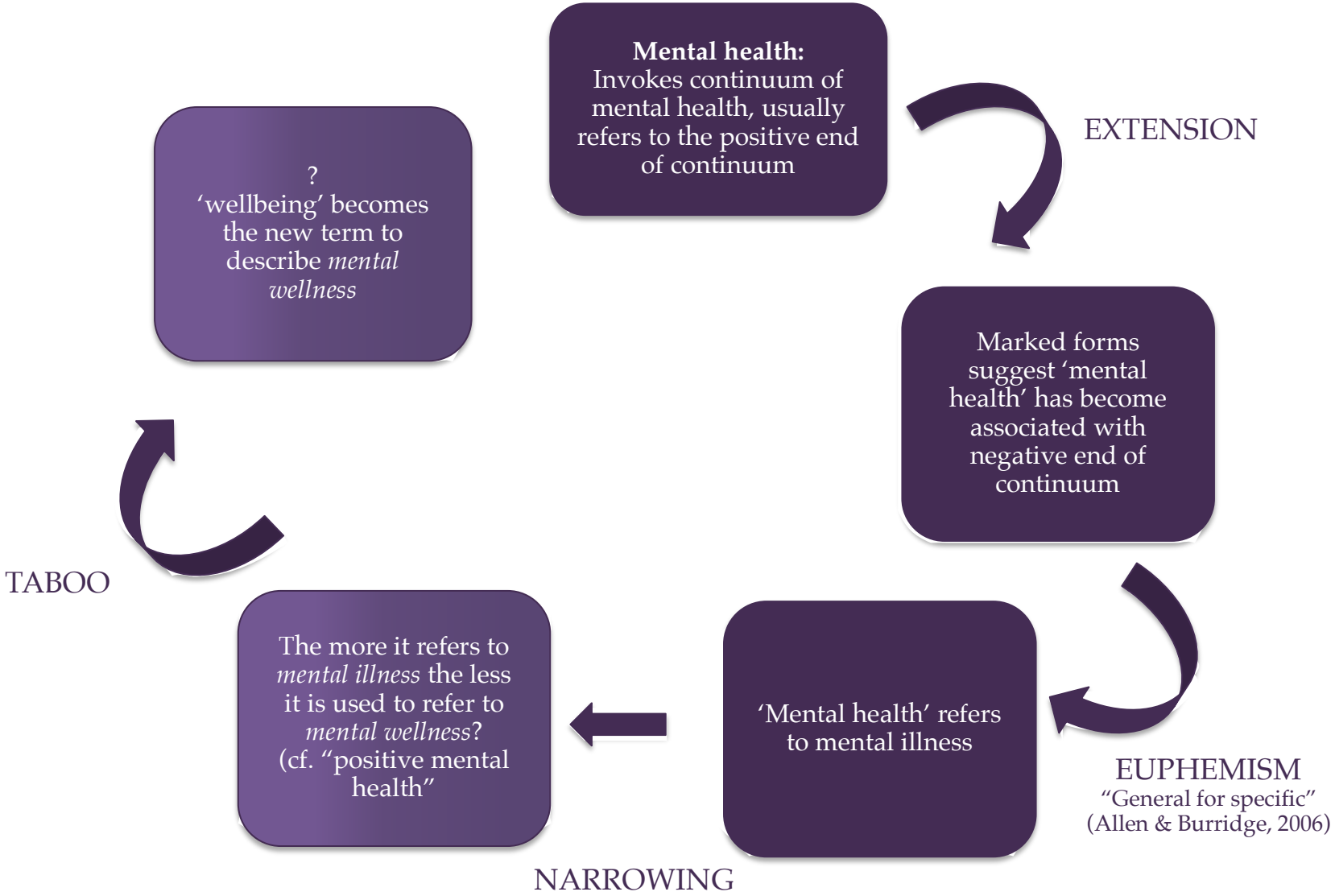


FIGURE 6.15. PREDICTED PROCESS OF SOCIALLY-MOTIVATED LANGUAGE CHANGE FOR 'MENTAL ILLNESS' (FROM MI 1984-2014 CORPUS) DATA)

So far in this chapter I have shown that definitions of 'mental health' published by major stakeholders in global mental health organisations suggest that 'mental health' is a term used to describe the scale of mental states, and also, in the absence of a positive term, to the state of *mental wellness.* I have argued that through a process of language change, specifically euphemism, the data suggests that 'mental health' now increasingly refers to *mental illness* in contexts where it does not refer specifically to organisations (e.g. trusts, foundations and charities). I argue that the reason 'mental health' is being used to refer to *mental illness* can be explained by using insights from pragmatics, because the shift in meaning from mental illness to mental health is the result of a socially-motivated language change (e.g. people do not refer to 'mental illness' in order to avoid the taboo associated with it). Due to this taboo avoidance strategy, the use of 'mental health' to refer to the *mental illness* constitutes a conversational implicature, because what is meant is more than what is said (Clark & Lucy, 1974; Grice, 1975; Levinson, 1983). The semanicization[69] (what might be termed the conventionalisation) of conversational implicature has been identified as a main mechanism for semantic change. For example, Sagi et al. (2009) write of pragmatic approaches to semantic language change that

> the main mechanism of semantic change is argued to be the semanticization of conversational implicatures, where conversational implicatures are a component of speaker meaning that arises from the interaction between what the speaker says and rational principles of communication
>
> (Sagi et al. 2009: 107)

Sagi et al's (2009) point is based on Levinson's account of pragmatic meaning (adopted by Traugott (2010) and Traugott & Dasher (2002), specifically his work on I-implicature which is a renovation of the Gricean approach (Grice 1975) to implicature,

[69] Traugott (2010) has previously referred to her work in theorizing semantic change as the the "semanticization of pragmatics".

and is particularly focused around a reformulation of Grice's (1975) Maxim of Quantity. Levinson's (2000) I-Principle (related to (I-Implicatures) refers to the informativeness of an utterance and can be used to explain the relationship between words within a semantic set, as well as why certain words are used over others within that set. Levinson (2000) posits a speaker's maxim for the I-Principle ("do not say more than is required" [Levinson, 2000: 100]) and a recipient's corollary ("what is generally said is stereotypically and specifically exemplified [Levinson, 2000: 100]). What this means is that speakers usually make their utterance informative enough for their recipients to understand the intended proposition, and that the recipient interprets the speaker's utterance as being designed in order to be informative enough. An example used by Atlas & Levinson (1981: 41) to exemplify an I-implicature is the utterance "John was reading a book" which implicates that John was reading a book that was not a dictionary (because when we read books we are not stereotypically reading a dictionary). As a result, "the hearer is licensed to derive the informationally enriching implicature that narrows down the non-specific predicate "book" by excluding dictionaries" (Carston, 1998: 194). In this example, what is said is more general than what is implicated. This fits with what we see in the data discussed in this chapter, e.g. why the more semantically general lexical item 'mental health' is used over the semantically more specific 'mental illness' to describe the concept of being *mentally unwell*. For my purpose here, the theory of I-Implicatures provides a useful means of explaining why 'mental health' may be used in the data to refer to *mental illness*. This is because 'mental health' is deemed informative enough to mean *mental illness*. That is, the term 'mental health' is taken to observe Levinson's (2000) I-Principle. Relatedly, I-implicatures explain this process because the way 'mental health' is being used in the data (i.e. to refer to mental illness) is not part of the formal semantic structure of the word; rather it arises from the specific discourse context and, as a result, it has taken on what Levinson calls "presumptive meaning" (2000) or a "preferred interpretation" (Chapman 2011: 101). Moreover, the ideas underlying the concept of I-Implicatures provide a useful means of conveying what I see as the

impetus for the language change outlined so far in this chapter – which is based on the conventionalisation of an I-Implicature. Specifically, 'mental health' has shifted meaning via the systematic use of 'mental health' to implicate *mental illness* (which happened as a result of the euphemistic language change that Traugott & Dasher (2002) identified in combination with speakers and hearers adhering to the I-Principle). In effect, 'mental health' has become conventionally associated with *mental illness* because a new definition of *mental illness* has been discursively constructed through the conventional use of 'mental health problem', etc. As a result, 'mental health', which traditionally referred to the continuum of *mental health* (comprising both good and bad) has taken on the presumptive meaning of referring to *mental illness*. It can be argued then, that the conventionalisation of 'mental health' to refer to *mental illness* in the data is due to an adherence to the I-principle (and relatedly, Grice's Maxim of Quantity [1975]) that stipulates that we do not say more than is necessary (i.e. if 'mental health' can refer to *mental illness*, there is no need to use the term 'mental illness' specifically).

Taken together, the preference for using a general term (e.g. 'mental health') over a specific one (e.g. 'mental illness'), and the preference for using the less taboo term fits with Traugott & Dasher's (2002) account of language change led by euphemism; i.e. the general expression 'mental health' is less taboo than 'mental illness' and therefore we can predict that the use of 'mental health' to mean *mental illness* will become conventionalised . This process is summarised in Figure 6.16.

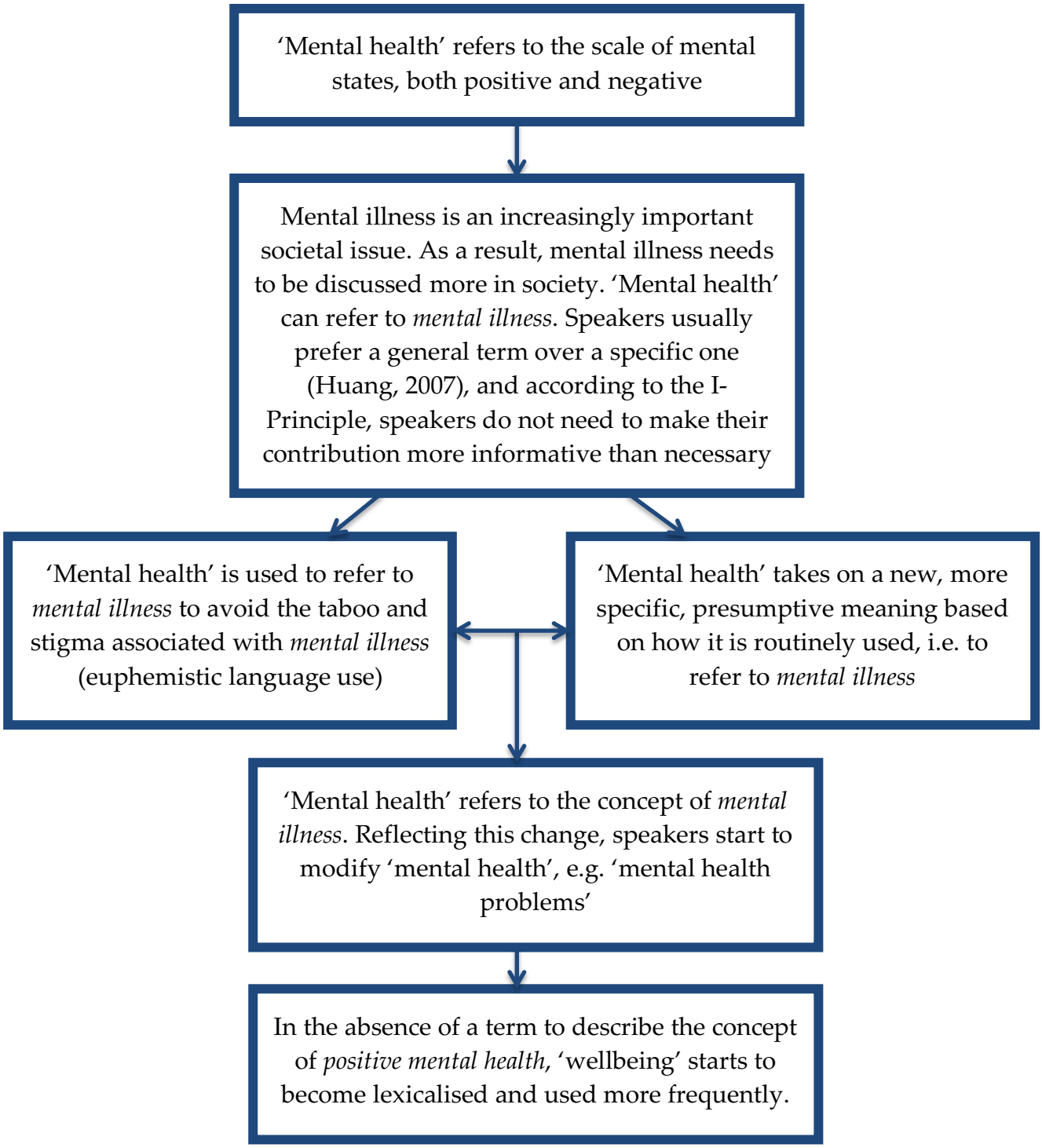


FIGURE 6.16. A PRAGMATIC ACCOUNT OF LANGUAGE CHANGE: MENTAL HEALTH TO MENTAL ILLNESS

Pragmatic accounts of language change provide a theoretically sound means of describing the change in usage within the data which shows that prior to the later

years contained in the corpus, mental health discourse has been underlexicalised (Fowler, 1977) (i.e. there is no term to denote *mental wellness* and *mental wellness* only). Arguably, the absence of a term conveying *mental wellness* has precluded discussion of *mental health* as a dynamic and changeable entity in the newspaper data contained in the corpus because *mental health* is most often represented as an issue of negative mental states and therefore discussion about maintaining *mental wellness* is rare. Moreover, because of the fact that the use of 'mental health' in the corpus can invoke *mental illness* but not vice-versa, the first concept invoked by 'mental health' will be *mental illness* which reiterates the view that there are only two states: *mentally ill* (which is referred to using the terms 'mental health'/'mental illness') or *not mentally ill*, which we don't have a term for. This binary view, which has possibly been compounded by the lack of terms denoting *mental wellness* only, is not an accurate representation of mental health, which can change at various points in a person's life. As previously stated, however, the increasing visibility of mental health and illness in society has meant that we are seeing lexicalisation in progress around this semantic domain. To return to the notion of I-implicatures, then, what we have in 'wellbeing' is a term that, unlike 'mental health' does not implicate mental illness.

Using the evidence presented since section 6.2, where I first outlined the mental health continuum based on the WHO definition of 'mental health', we can now revisit the continuum of mental health enriched by our understanding of how the data suggests these concepts, i.e. *mental illness, mental health, mental health problems/issues/conditions,* and *wellbeing* occur.

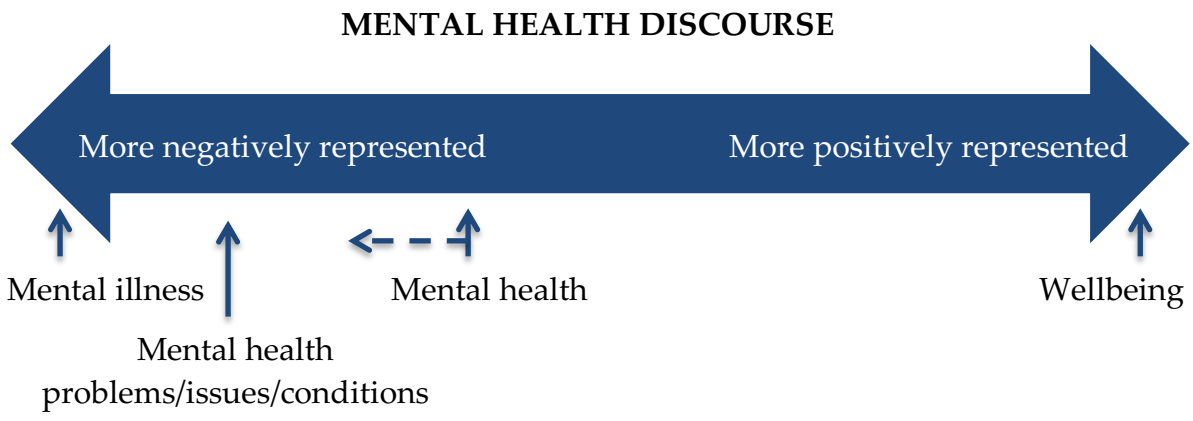


FIGURE 6.17. THE CONTINUUM OF MENTAL HEALTH DISCOURSE BASED ON TERMS FROM 2013

Figure 6.17 shows that during the period examined, 'mental health' was not always viewed as a term to describe the continuum of states between *mental health* and *mental illness*. Rather it is used as a neutral term to describe legislation or official bodies concerned with mental illness. Furthermore, it is used in an increasing number of cases to refer to negative mental states. Taken together, these factors suggest that within newspaper discourse across the years sampled, 'mental health' has taken on a new discursive meaning, distinct from that of the official discourse of mental health outlined by WHO, which suggests 'mental health' refers to the scale of mental states. This discursive meaning of 'mental health' emergent in the data conveys *mental illness*.

If we compare the contemporary picture of mental health discourse to that of just over 30 years ago (1984), we see that it is not at all surprising that a new positive term has started to emerge. The reason for this is because 'mental health' is used solely to refer to legislation, specifically the Mental Health Act and any other mental states were covered by the label 'mental illness'. A possible reason for the fact that 'mental health' was used solely to refer to legislation is that *mental wellness* or the maintenance of it was not discussed in the mainstream media. The only state on the continuum that was discussed was *mental illness* because the absence of *mental illness* was not viewed

as *mental health* or *wellbeing* but as, presumably, 'normal' which did not require a term. What we can conclude from this, then, is that over the time period the visibility of issues to do with mental health and illness has increased, which has meant that new terms for this in this discourse have emerged, as would be the case in the lexicalisation of any semantic domain. What this means in real-world terms is that whilst *mental illness* is still the focus of press reports by a significant margin, *mental wellness* is being discussed more and more. This is a hopeful change in the discourse around mental illness as it works to normalise the idea that mental health can fluctuate, that mental health and illness are not binary, that speaking about experiencing a range of mental states is societally acceptable, and relatedly that the absence of mental illness is not necessarily 'normal'.

Further to the changes perceived in the data, an interesting observation I have made since the start of 2014 (the corpus only covers January 2014) that attests to 'wellbeing's' role as a new term to denote *mental wellness*, can be seen on the WHO website which was updated in 2014. The definition of 'mental health' listed there is:

> Mental health is defined as a state of well-being in which every individual realizes his or her own potential, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to her or his community.
>
> (WHO 2014)

Furthermore, the Mind website in the year of writing (2019) defines *wellbeing* as "how you are feeling and how well you can cope with day-to-day life" (Mind, 2019c). This definition of *wellbeing* shares some of the semantic content of the WHO definition of health, which is "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity." (WHO, 2019), thus supporting the argument that since 2014 'wellbeing' has gained traction as a term meaning *mental wellness.*

6.7. Conclusion

In this chapter, I have argued, using evidence from language in use, that the labels for concepts within the semantic domain of MENTAL HEALTH AND ILLNESS are distinct and shifting. This study is the first to my knowledge that has used large quantities of authentic language data to track semantic change within the semantic domain of MENTAL HEALTH AND ILLNESS. In doing this, I have laid the groundwork for Chapters 7 and 8, in which I explore the participants (through naming analysis) and processes associated with mental health and illness.

The findings from the analysis conducted in this chapter relate to the point I made at the beginning of this chapter concerning the terms used to collect data (the Day & Page [1986] and Wahl et al. [2002] examples). I have shown throughout the analysis that the terms 'mental health' and 'mental illness' are distinct enough to potentially bias data collected using one term over another term (e.g. 'mental health' often refers to *mental illness*). For this reason, the findings reported in this chapter have wider implications for the methodological best practice of studies whose aims are to model the semantic domain of MENTAL HEALTH AND ILLNESS. Moreover, to the best of my knowledge, no research has been conducted into the terms 'mental health' and 'mental illness', despite the fact that much research has been conducted into the terms for individual illnesses, e.g. schizophrenia. This lack of research suggests that the terms may be taken for granted. Through the analysis conducted in this chapter, then, I have shown that the meanings behind 'mental illness' and 'mental health' cannot be taken for granted.

In summary, in this chapter I have shown that the terms 'mental health' and 'mental illness' have increased and continue to increase across the time period studied. I have suggested that this indicates that *mental health* and *mental illness* are increasing in both importance and visibility in UK society. Through an exploration of 'mental health' and 'mental illness', it appears to be the case in the MI 1984-2014 corpus that 'mental illness' is becoming the dispreferred term to refer to negative

mental states, while 'mental health' is becoming more common and subsuming some of the sense of *mental illness*. This finding is supported by the increasing use of modifiers of 'mental health' that carry negative meanings, such as 'mental health problems' and 'mental health disorders'. Moreover, the MI 1984-2014 corpus contains more articles reporting on 'mental illness' than 'mental health' due to the sampling frame used to collect the data. Due to the data collection procedure adopted in the creation of the MI 1984-2014 corpus, then, it would be sensible to assume that there would be a higher proportion of instances of 'mental illness' than 'mental health' but this does not bear out in the data. Despite the fact that the corpus contains a higher proportion of articles reporting on mental illness, the analysis conducted in this chapter has shown that 'mental health' is a more frequent term than 'mental illness', and that 'mental health' is rising. This provides further evidence for my argument that 'mental health' has taken on some of the sense of 'mental illness' in this discursive context[70]. I have also argued that if 'mental health' is now being used to refer to negative mental states, then, just as 'mental health problem' emerged to convey *mental illness* (negative meaning the other end of the health continuum), it could be expected that a new term will emerge in order to convey *mental wellness.* One candidate term to fill this lexical gap is 'wellbeing'. Evidence from the MI 1984-2014 corpus shows that the use of 'wellbeing' is increasing considerably, despite the term appearing to be relatively new within the semantic domain of MENTAL HEALTH AND ILLNESS. Moreover, I have provided evidence to show that 'wellbeing' is undergoing a process of lexicalisation, which offers further evidence for the claim that 'wellbeing' is becoming an established term in this semantic domain. The diachronic development of this term is something that warrants further exploration and which unfortunately requires more contemporary data than is contained in the MI 1984-2014. However, it is possible

[70] Without data to compare the frequency of mental health and mental illness in other discourse types, e.g. in clinical settings, it is impossible to test whether this is true of other settings. My intuition is that 'mental illness' would still be the preferred term to refer to mental ill health by medical experts, although anti-stigma initiatives in the UK refer to 'mental health' and they are supported by mental health campaigners and medical experts.

at this early stage of the term's development to make hypotheses about 'wellbeing'. One such hypothesis is that it could be a useful alternative in mental health awareness-raising initiatives, as the term has not yet been linked to the word 'stigma'. Moreover, based on qualitative observations from the year of writing, it does appear that the term 'wellbeing' has developed significantly since 2014, giving an early suggestion that the semantic processes outlined in this chapter have borne out, thus demonstrating the utility of the linguistic analysis of societal discourses.

Using methods from pragmatic theory, I have demonstrated that the diachronic change observed in the analysis is entirely consistent with socially-motivated, pragmatic accounts of language change, wherein euphemism (e.g. using 'mental health' to refer to *mental illness*) provides a catalyst for shifting semantic meaning. Further to this, using the concept of I-Implicatures (Levinson, 2000), I have also shown how 'mental health', came to refer to *mental illness*. This happened via a process of implicature relating to the informativeness of utterances, specifically the process by which 'mental health' took on the presumptive or preferred meaning of 'mental illness'. This chapter, then, has synthesised a range of evidence to give an overall picture of the contemporary discourse of mental health and illness contained in the corpus. In the next chapter I explore the naming practices used in the corpus to refer to people with mental illness.

7. Named, labelled and referred to: people with mental illnesses in the MI 1984-2014 corpus

7.1. Introduction

In Chapter 6, I showed that the terms used to describe mental illness and mental health have shifted and are continuing to shift. I focused my analysis of these changing labels on the semantic and pragmatic content of 'mental health' and 'mental illness'. Specifically, I argued in Chapter 6 that the changing meanings of 'mental illness' and 'mental health' were an example of pragmatic-led language change guided by euphemistic language use. In this chapter I explore the labels used to describe mental illness in more detail; however, in this chapter I focus on the way that labels can encode ideology (as opposed to what they mean semantically). Specifically, in this chapter I explore the labels that are used by journalists to describe people with mental illness in more detail. My analysis in this chapter, then, addresses two of the research questions listed in the introduction to this thesis. These are:

1. What linguistic strategies are used to name, label and describe people with mental illness?
   1.1. To what extent is person-first language present in the MI 1984-2014 corpus?
2. What themes are present in the corpus for referring to people with mental illness?

As previously stated, this chapter is concerned with the ideological content associated with using certain labels to name and refer to people and entities in the world. The reason for conducting naming analysis is because, as I described in Chapter 2, it allows for the exploration of how participants or entities are "packaged-up" (Jeffries, 2010: 19). I argued in Chapter 4, in line with previous research on naming (Clark, 1992;

Jeffries 2010; *inter alia*), that naming analysis constitutes a linguistic method for analysing ideology in texts. The reason that naming analysis is particularly useful when applied to newspaper discourse is that, as Bell (1994) states, "journalists do not write articles. They write stories. A story has structure, direction, point, viewpoint" (Bell, 1994: 100). Put simply, what may appear to be an accurate depiction of an event has often undergone a process of what we might term 'storification'; i.e. the elements of the story that are the most attention-grabbing or shocking are foregrounded (sometimes literally via headlines) whilst other elements (which may contextualise the event) are backgrounded. This storification process is useful for the analyst because it makes salient those elements of the reports that the journalist deems attention-grabbing or shocking, which in turn allows for the linguistic analysis of what constitutes newsworthiness and news values (Bednarek, 2006; Fowler, 1991; Richardson, 2007), or "values that *exist in and are constructed through* discourse" (Bednarek & Caple, 2014: 135, original emphasis). More interesting, however, for my purposes is how those attention-grabbing elements (for example, salient social actors and salient circumstances) are named in order to be attention-grabbing or shocking. For example, a woman with a mental illness may be described as 'bipolar mother' despite the fact that her maternal role is not pertinent to the story. Why is it relevant that the journalist mention that she is a mother? Moreover, what does naming the woman in this way indicate about society's views of motherhood as well as society's views on bipolar disorder, and how does this relate to the newsworthiness of the story? The answer to these questions, I argue, is to be found in the tendency of the press to put people into categories or socially-constructed groups. Fowler (1991) describes such groups as being an "instrument for handling discrimination, for sorting unequally" (Fowler, 1991: 94). To return to the 'bipolar mother' example, the story is arguably newsworthy because the woman belongs to the category 'people with mental illness' who are of a particular 'group' and who are routinely represented in the press as being criminal, violent and unpredictable (Bowen, 2016; Whitley & Berry, 2013; Paterson, 2006, *inter alia*). As a result, her belonging in this category clashes with

her belonging to the other 'group' – that of 'mothers' (e.g. women who are stable, dependable, etc.). Presenting the woman as 'bipolar mother' then is one way that journalists can create covert associations that constitute damaging and baseless ideological content (e.g. that a diagnosis of bipolar disorder affects the ability of a woman to be a good mother according to common associations of what motherhood constitutes). Naming, then, is both a referential device, in that it allows for texts to point at people in the world (who may be mothers and may have a mental illness), and a narrative device, in that it constructs them as a particular character in a story (the bad mother, the neglectful mother, the unpredictable mother). Naming practices, then, are the link between the complex and multifaceted people and things in the world and the story constructed in texts (a story that, for our purposes here, has been constructed by journalists). This is because categorisation is necessarily a simplification of a person or a thing (Rosch, 1975). As Fowler (1991) writes

> Having established a person as an example of a type, our relationship with the person is simplified: we think about the person in terms of the qualities which we attribute to the category already preexisting in our minds […] the category may harden into a stereotype, an extremely simplified mental model which fails to see individual features, only the values that are believed to be appropriate to the type
>
> (Fowler, 1991: 92)

In addition to the role that language plays in creating and perpetuating stereotypes (for example presenting a group of people in an unfavorable and untrue way), the way that the press chooses to categorise and name people has implications for the people categorised and named. Consider, for example, the relationship between personal identity (how we identify ourselves) and our social identity (how others identify us) discussed by Ryan et al. (2009):

> we are not solely who we think we are, but also who others believe us to be; we come to learn about ourselves through the reactions of others […] Consequently, we are all in danger of being socially positioned in unfavorable ways; that is, to having our desired projected selves rejected and other undesirable selves thrust upon us
>
> (Ryan et al. 2009: 146)

Ryan et al. (2009) essentially argue that the way we are represented (for our purposes here, represented *in language*) affects not just how others perceive us but also how we perceive ourselves; i.e. we start to internalise negative stereotypes about ourselves. This point brings us back to some of the findings of previous research reported in Chapter 2 on self-stigma, such as that conducted by Schomerus et al. (2012) who found that self-stigma resulted in people not recognising their symptoms as being related to mental illness, and also resulted in people not engaging with mental health professionals (recall that self-stigma relates to a person devaluing themselves as a result of stigmatising attitudes about them projected by others (Corrigan et al., 2010; Goffman, 1963)). For this reason, mental illness is an important topic matter to explore in terms of naming strategies because the ways in which people with mental illness are named can not only present a simplified and stereotypical view of people with mental illness in public life generally, but can also have real-world, personal health consequences for individuals with mental illnesses.

In the next section, I discuss prescribed linguistic forms for referring to people with mental illness. In Section 7.2, I detail the analytical method deployed in this chapter before I report the results of the naming analysis in section 7.3. In section 7.4, I discuss my findings and conclude.

7.1.1. A linguistic prescription for mental illness stigma?

In addition to the reasons I have given above for why naming analysis is of interest for analysing ideology in language, the data itself presents some interesting ideological issues. The reason for this is that during the period represented by the MI 1984-2014 corpus there have been preferred and prescribed ways to refer to people with mental illness. For example, in the 1990s organisations such as the American Psychiatric Association (hereafter APA) advocated for 'person-first language' (Granello & Gibbs, 2016; Halmari, 2011). Person-first or people-first language refers to when a particular diagnosis (sometimes called a designation) follows the head noun in a prepositional phrase or a relative clause (e.g. *a person with mental illness* or *a person who is experiencing mental illness* vs. *a mentally ill person*). As Granello & Gibbs (2016) write, "Person-first language was offered as a mechanism to separate the identity of the individual from any clinical diagnosis, disability, or chronic condition." (Granello & Gibbs, 2016: 31). Due to the fact that other forms by contrast do not distance the identity of the individual from the diagnosis, some researchers refer to non-person-first language as 'identity-first' (Gernsbacher, 2017).

Since the 1990s, prescribed forms such as person-first language have become more widely adopted and are promoted by anti-stigma initiatives such as the *Time to Change* campaign launched in 2007 by the UK mental health charities *Mind* and *Rethink Mental Illness* and funded by the UK Department of Health and Social Care. On their website, *Time to Change* list several naming strategies that they advise journalists to avoid using including many identity-first forms (although *Time to Change* do not use that term). They write that "Choosing the right language to describe people with mental health problems is important. Using inaccurate terms can reinforce stereotypes and stigma" (*Time to Change*, 2019). In their guidelines, *Time to Change* advise journalists to avoid using derogatory terms formed by shortening mental illnesses such as 'a psycho' and 'a schizo', and identity-first terms such as 'a schizophrenic' 'the mentally ill' and 'victims'. In their guidelines, *Time to Change* promote the use of

person-first terms such as 'a person who has experienced/is experiencing depression' and unmodified nouns that are not, by definition, related to mental illness such as 'patients', 'service users' and 'clients' (*Time to Change*, 2019). *Time to Change* list several reasons why certain linguistic forms should be avoided. For example, avoiding words such as 'maniac' and 'mad' is advised because "These words are usually linked to dangerousness or strange behaviour" (*Time to Change*, 2019). Despite giving these reasons (and the fact that to a speaker of English the associations *Time to Change* mention seem intuitively to be correct), *Time to Change* do not cite any linguistic research into the link between the terms listed and those associations. For this reason, I will return to the issue of person-first and identity-first language in relation to naming strategies used in the MI 1984-2014 corpus in Section 7.3.1. In addition to exploring the naming strategies used to refer to people with mental illness mandated by *Time to Change* in this chapter, I will revisit the prescribed forms promoted by *Time to Change* in relation to transitivity processes in Chapter 8.

To briefly return to the linguistic features of newspaper discourse discussed in Section 7.1 in the context of person-first language, previous research has shown that newspaper articles are prone to storification and simplification (e.g. through stereotyping) in order to sensationalise events. On the one hand, newspapers constitute a major source for public information about mental illness (Nawková et al., 2012: 22); on the other hand, the stories need to be succinct and engaging. The purpose of newspaper articles, then, is to inform but also to entertain. In addition to its entertainment function, newspaper discourse utilises particular structural features into which person-first language may not be easily incorporated; for example, newspaper headlines that are intended to be shocking or eye-catching. Halmari (2011) writes that person-first language (which she refers to in this extract as politically correct, or 'PC' language) is not always conducive to the brevity associated with headlines:

> Newspaper headlines are notorious in doing away with function words in order to save space. The 'people first' language, by contrast, is not interested in saving space; it is notorious for its circumlocutionary style. 'People first' circumlocutions frequently necessitate the use of function words such as prepositions […] and relative pronouns. […] headlines, hence, have a natural need to resort to non-PC language, if it were only to save space.
>
> (Halmari, 2011: 837)

As Halmari (2011) alludes to, the naming strategies used to refer to people with mental illness raise interesting questions for the analyst because, unlike other topics, the semantic field of mental illness has prescribed linguistic forms; however, newspaper discourse has fairly rigid textual features (both in terms of linguistic structure and genre norms to do with engaging writing). For this reason, the analysis of the naming strategies in this chapter is novel in that (I) it provides new systematic research into practices used to refer to people with mental illness and mental illnesses themselves, and (ii) because it also explores in more detail the linguistic basis of person-first language in relation to mental illness (see Halmari (2011) for a linguistic analysis of person-first language in general). I return to person-first language and other prescribed linguistic forms in Chapter 10.

In the next section, I outline the analytical method used in this chapter.

7.2. Analytical method

As explained in the introduction, naming analysis is concerned with how people and entities are labelled in texts, and what ideological effect these labels may have. In order to collect instances of labels in the corpus so as to conduct naming analysis, I collected headlines from the corpus using stratified sampling by year. To sample from the corpus, I collected the first and last three headlines from each year that pertained to people with mental illness or a mental illness (I refer to this sample as 'the headline

sample' in the analysis). When sampling from the headlines, I did not include articles on autism or articles on issues only tangentially related to mental illness. The reason for collecting headlines, as opposed to collecting the labels used in whole articles was that, as Page (2003) (in line with Bell, 1991 and White, 2000) argues, "the headline occupies a position of textual and evaluative prominence in the news report (Bell, 1991; White, 2000) [and] it might be predicted that […] these fragments would be of particular importance and indicative of the emphases assigned to the identities associated with the various naming choices" (Page, 2003: 563-64). Furthermore, as Conboy (2007: 15) writes, "headlines are in themselves a distinctive contribution to the news values of a newspaper in the ways that their syntactic structure can give patterned evidence of stylistic preference". Headlines, then, are a useful starting point from which to commence naming analysis of newspaper articles because (i) they are indicative of who journalists deem to be newsworthy participants (i.e. people) and (ii) they often provide a summary of the article (i.e. the circumstances the person is described as being involved in or related to). Furthermore, headlines provide "a lens on, stance towards or angle on the rest of the story" (Caple, 2013: 276). Furthermore, although I am not analysing the multimodal aspects of newspaper articles in this thesis, it is worth noting that headlines are typically printed in larger font, and in some cases capitalised. As a result, headlines are visually prominent in addition to being textually prominent. By the terms of foregrounding theory (e.g. van Peer 1986) then, headlines are more likely to be remembered by readers than the main body of the article. This is another reason why headlines provide rich data to the critical linguist interesting in analysing ideology in texts: headlines are remembered in a way that the specific circumstances of the article are not (take for example, the 'Bonkers Bruno' headline example I referred to in Chapter 1).

Once I had collected the headlines, I then qualitatively analysed each headline for salient naming strategies (e.g. '*the mentally ill*') and key themes in the way that people with mental illness were referred to (e.g. referring to people with mental illness

as quantities such as *'1 in 4'*). An example of the categorisation I used is shown in Table 7.1 below:

Referent type	Instance of naming
Mentally ill	Mentally ill The mentally ill Mentally ill homeless Mentally ill offenders Mentally ill people
patient	Bulimia patient Cheltenham patients Patient Patients PATIENTS WHO SUFFERED TORMENT second patient free to kill victims

TABLE 7.1 EXAMPLE OF CATEGORISATION TABLE FOR NAMING PRACTICES BY REFERENT TYPE

In addition to the headline analysis, I also used stratified random sampling by year to explore the salient naming practices and themes in references to people with mental illness in whole articles. I did this by assigning a random number to each article and randomly selecting three whole articles from each year (I refer to this sample as the 'sample by year' in the analysis section). For this sample, I categorised each referent in the data according to four categories: (i) 'person with mental illness', (ii) 'medical expert', (iii) 'other' and (iv) 'terms for mental illness''. Due to the fact that this sample included the main body of newspaper articles, I was also able to categorise reference chains (e.g. *'John'*, *'Mr Smith'*, *'John the schizophrenic'*) in more detail than the headline sample allowed (because headlines are shorter in length and therefore are less likely to feature anaphoric or cataphoric reference). This sample, then, allowed me to explore in detail whether the naming practices identified in the headlines were also a salient feature of other articles in the corpus, as well as to explore anaphoric reference in more detail. The salient naming practices identified in the two samples then formed the basis for corpus analysis wherever relevant; for example, to explore the frequency of

certain naming conventions over the time period covered by the MI 1984-2014, or to explore the collocates of certain labels attributed to people with mental illness. In the next section, I present my analysis.

7.3. Analysis

As previously stated, the analysis in this section is concerned with what naming strategies are present in the data, and what ideological implications these naming strategies may have. For this reason, the analysis section is presented in three sections, each addressing a separate facet of naming practices in the corpus. In the first of these sections, Section 7.3.1, I explore the frequency of person-first vs. identity first forms in the data samples and in the MI 1984-2014 corpus because these forms of labelling have previously been reported as contributing to negative ideologies about mental illness. The second section, 7.3.2, explores in more detail the ways that the press name people with mental illness. The third section, 7.3.3, reports on a salient theme identified in the data for referring to people with mental illness as quantities. I then discuss my findings and conclude this chapter in Section 7.4.

7.3.1. 'A person experiencing a mental illness' or 'a mentally ill person'? Exploring person-first and identity-first labels in the MI 1984-2014 corpus

In Section 7.1.1 I introduced person-first and identity-first language as two ways in which a person may be described in relation to their illness. Person-first terms typically relate to postmodified nouns, e.g. a person *with mental illness*, whereas identity-first forms relate to premodified nouns, e.g. a *mentally ill* person. The theory behind person-first language is, as Granello & Gibbs (2016) state, "ideologically grounded in the principle of linguistic relativity (popularly known as the Sapir–Whorf hypothesis), which states that language shapes perceptions of the world and significantly influences cognitive processes" (2016: 31). The version of linguistic

relativity subscribed to by Granello & Gibbs (2016) leads to the suggestion that changing the way we talk about people with mental illness will change the way we think about people with mental illness. If we accept that this is the case, as a result of removing stigmatising language we may well predict that a reduction in stigmatising attitudes towards people with mental illness. Given that my aim in this section is to explore the frequency of person-first language in reference to naming and not to analyse the linguistic basis of prescriptive forms, I will not problematise the theory underpinning person-first language in this section. I will, however, return to the linguistic theory underpinning prescribed linguistic forms such person-first language in the conclusion of this thesis in Chapter 10.

In addition to introducing person-first language in Section 7.1.1, I also discussed how person-first language is the linguistic form preferred by anti-stigma initiatives in the UK such as the *Time to Change* campaign. In addition to listing several identity-first labels to avoid in their media guidelines, *Time to Change* also state that terms that label people with mental illness as their illness, such as 'a schizophrenic', should also be avoided due to the fact that "People are more than their illness, it doesn't define them." (*Time to Change*, 2019). In this section, I provide an initial analysis of the sample headlines and the MI 1984-2014 corpus to explore the frequency and distribution of preferred forms and terms deemed to be problematic for naming people with mental illness.

To establish whether person-first forms were present in the corpus, I explored the sample headlines. Of the 186 headlines in the sample, only three featured person-first language and two of those three featured other descriptions deemed problematic by *Time to Change*. Table 7.2 shows the headlines:

Date	Headline	Publication/ Author
Feb 18 1988	Thursday Women (Life Forces): Seeing the light - **People suffering from winter blues** are being treated with artificial sun rays	Guardian, By JILLIE COLLINGS
Jan 2 1999	Suicidal acne drug 'victims' are to sue for compensation; LANDMARK CASE LOOMS AS **PATIENTS WHO SUFFERED TORMENT** WIN LEGAL AID	Daily Mail, By Emily Wilson
Dec 29 2008	Diet-battle teens get place of own; HEALTH: New centre offers specialist help for **children with eating disorders**	Birmingham Evening Mail, By Alison Dayani

TABLE 7.2. HEADLINES INCLUDING PERSON-FIRST LANGUAGE IN THE HEADLINE SAMPLE

In the first two headlines (taken from 1988 and 1999), the person-first forms "people suffering from winter blues" and "patients who suffered torment" (where torment refers to suicidal thoughts) are both used in conjunction with the verb 'suffer', which *Time to Change* state is a word to be avoided (I conduct an analysis of 'suffering' in Chapter 8). The third example features person-first language, "children with eating disorders", and while it does not include the verb 'suffer', it does include a summary of the story in which the journalist has chosen to represent the children with eating disorders using the hyperbolic label "diet-battle teens". The fact that the person-first form was applied to child referents is in line with Halmari's (2011: 828) finding that "postmodification is reserved for children or non-criminal adults". The low frequency of person-first language in the headlines is perhaps unsurprising given that, as previously stated, headlines are necessarily short; therefore, circumlocutory language such as person-first language may be avoided. For this reason, I will now report the frequency of person-first forms in the MI 1984-2014 corpus as a whole.

In order to assess how often person-first forms were used in the MI 1984-2014 corpus in contrast to other forms, I selected labels for mental illnesses that could be used to identify the individual as well as those which could be used in a person-first frame (e.g. 'a schizophrenic' and 'a person/people with schizophrenia'). Figure 7.1

shows the relative frequency for each variant (i.e. the -*ic* suffix variant, the *person with* variant, and the *people with* variant).

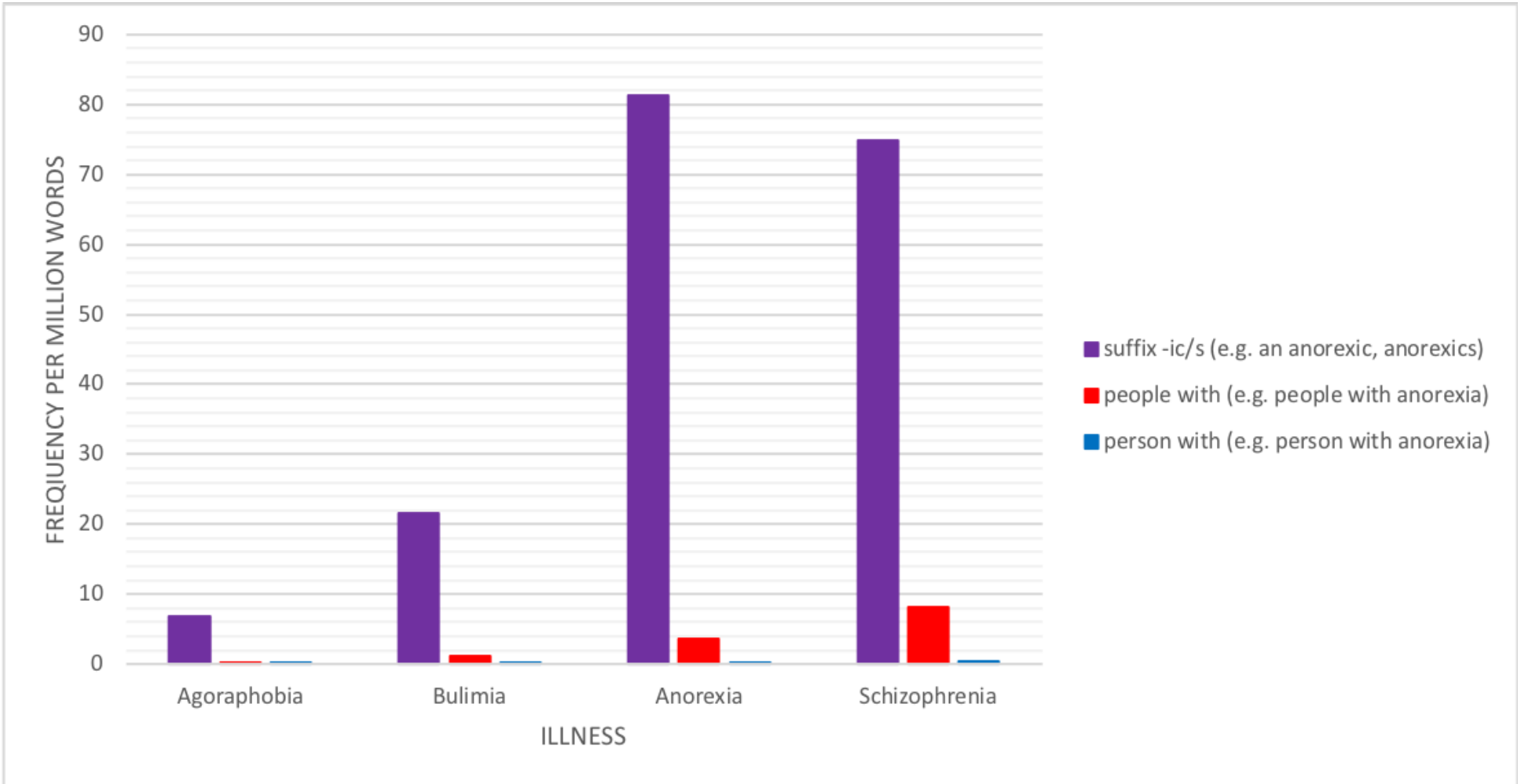


FIGURE 7.1. FREQUENCY OF IDENTITY-FIRST VS. PERSON-FIRST FORMS FOR AGORAPHOBIA, BULIMIA, ANOREXIA AND SCHIZOPHRENIA)

Figure 7.1. shows that the identity-first form is much more common in the corpus overall than person-first forms. Moreover, of the person-first forms, the plural variant is much more common overall that the singular variant. This is indicative of another theme in the naming of people with mental illness in the corpus, which is to refer to people with mental illness as quantities (and therefore plural forms are more common than singular forms). I will discuss this theme in more detail in Section 7.3.3. In order to explore the trend in use of person-first forms in the MI 1984-2014 corpus over time, I plotted the relative frequency of "person with" and "people with" over the 31-year period covered by the corpus. The reason for looking at the diachronic distribution of person-first forms was to explore the hypothesis that despite the fairly low frequency of person-first forms in relation to identity-first forms (as shown in Figure 7.1), the overall trend for person-first forms would be positive; i.e. due to the press adhering to prescribed forms, there would be an increase in person-first language over time. The reason for this is that the findings I reported in Chapter 6 also give indirect support for this hypothesis. These findings showed that there was an overall preference in the corpus for people to refer to mental illness via euphemistic reference such as 'mental health' and 'mental health problems'. As a result, those findings provide further support for the hypothesis that language over time becomes more euphemistic (i.e. increasing use of person-first language). Figure 7.2 shows the frequency of [person with (a) mental], [people with (a) mental] and [people/person with schizophrenia]

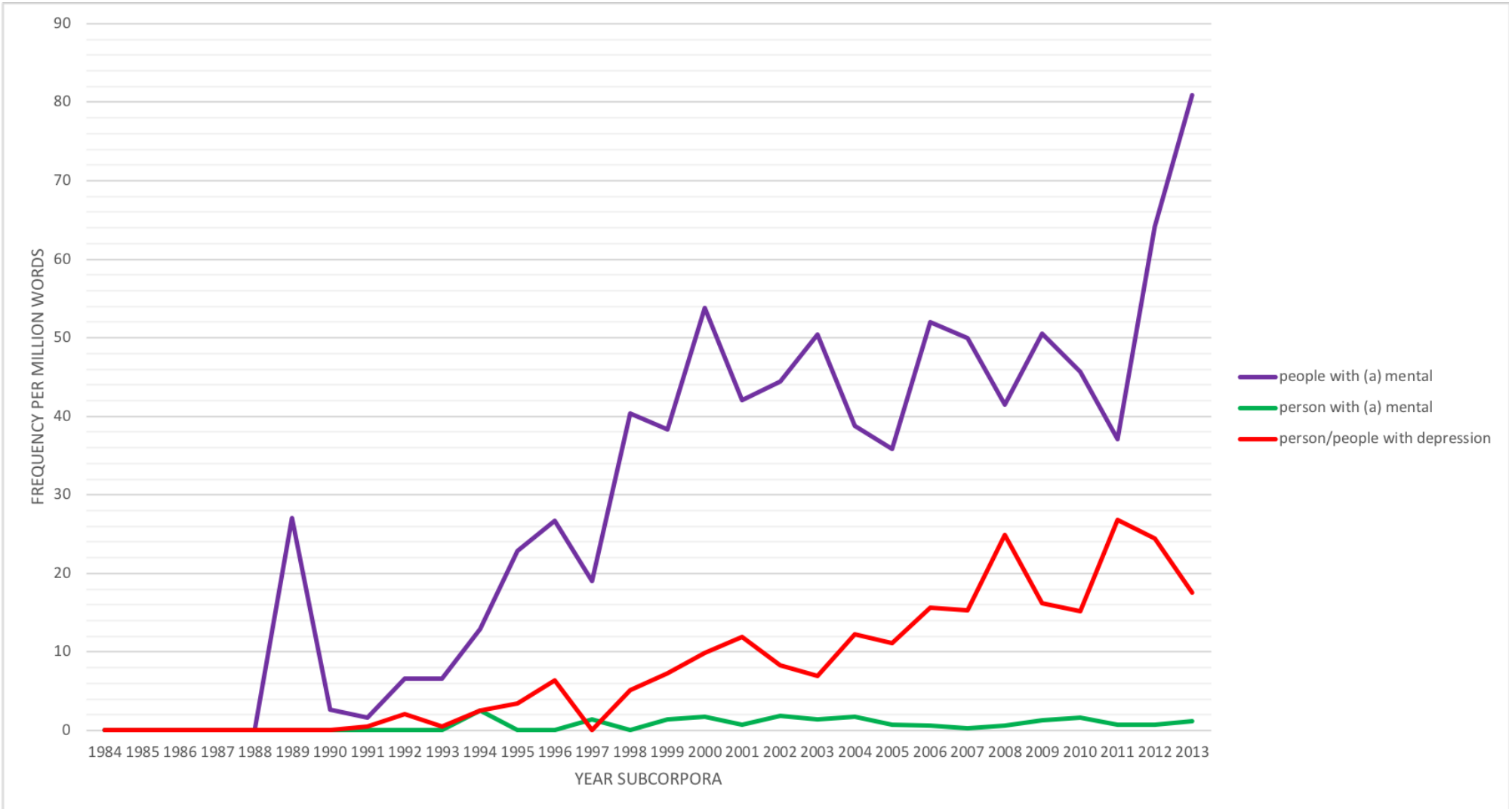


FIGURE 7.2. FREQUENCY OF PERSON-FIRST FORMS OVER TIME IN THE MI 1984-2014

Figure 7.2. reveals that despite the low frequency of person-first forms over the time period, the frequency overall is increasing. This is in support of the hypothesis that, in general, euphemistic forms for referring to mental illness are increasing. The increased usage of person-first language over time is in line with the research findings of Barnish (2014), who found that in a collection of research articles on disability, "There was a significant increase in the use of person-first language […] over the time period 1994-2013, although it remained a minority usage" (Barnish, 2014: 505). The increased usage in person-first forms shown in Figure 7.2 start to rise around the late 1990s, which is in accordance with Granello & Gibbs' (1990s) description of person-first language as being adopted by various organisations in this period. Furthermore, politically correct (or PC) language (a category that person-first forms fall into) has previously also been identified by linguists as a 1990s phenomenon. For example, Cameron (1995) refers to political correctness as "the 1990s *zeitgeist*" (1995: 116).

In addition to including person-first forms for mental illness generally (collected using the search term [person/people with (a) mental], I also included person-first references to the specific condition of depression. The reason for doing this was to assess whether person-first language was also increasing in reference to specific illnesses. I used 'depression' as a search term over the other illnesses contained in the corpus because depression is the most common illness referred to in the MI 1984-2014 corpus. In addition to providing further evidence that person-first forms are rising over time, including the depression condition (i.e. [person/people with depression] also supports the finding I have mentioned previously in this section that singular referents are less commonly referred to than plural referents; i.e. there are more instances of (plural) people with schizophrenia than there are references to the singular '[person with (a) mental]'. This is noteworthy because we might have expected references to mental illness generally (e.g. a person with mental illness) to be more common than references to specific mental illness (e.g. a person with depression) according to the pragmatic principles outlined in Chapter 6, whereby euphemistic

language is typically more general than specific (e.g. referring to a specific mental illness as 'mental health'. Recall that Allen & Burridge (2006) identified euphemisms as being created by a 'general-for-specific' substitution. What my finding means is that in the corpus, journalists are more likely to refer to plural 'people' than singular 'person' in person-first forms, even when they are referring to specific illnesses. I will discuss the tendency in the corpus to refer to people in groups or quantities in Section 7.3.3.

So far in this section I have explored the use of person-first language and looked at the frequency of person-first forms in the headline sample and in the MI 1984-2014 corpus as a whole. Now I will look in more detail at the context of person-first and identity-first forms in the corpus to explore whether person-first forms occur in more positive or favourable contexts to people with mental illness as the theory behind person-first language would suggest. Table 7.3 shows the top three randomly generated concordances for the person-first structure 'people with mental illness' in the MI 1984-2014 corpus.

Person-first examples 'people with mental illness' (Raw freq. 797, Rel. freq. 13.61pmw)		
This film has its facts totally wrong and turns	**people with mental illness**	into figures of fun
"It is a fact that there are many more murders by so-called normal people than by	**people with mental illness**	
its vision is to get people to realise the implications of placing stereotype and stigma labels on	**people with mental illness**	

TABLE 7.3. CONCORDANCES FOR 'PEOPLE WITH MENTAL ILLNESS' TAKEN FROM THE MI 1984-2014 CORPUS

Table 7.3 shows that in all of the concordance lines featuring person-first language, the surrounding context features metacomment on the representation of people with mental illness in the media. For example, the first two concordances discuss the misrepresentation of people with mental illness in the media. More interestingly still, the third concordance line discusses the labels used to describe people with mental illness, particularly in reference to labels that stigmatise individuals. Further analysis of other examples of 'people with mental illness' in the corpus also reveal that metacomment (particularly in relation to stereotypes of people with mental illness) is a common feature of person-first forms. Further evidence for the finding that person-first language occurs in contexts in which people are challenging stigma around people with mental illness is that the statistically significant collocates of 'people with mental illness' include the words 'myth', 'discrimination', 'stigmatised', 'belittle', 'stigma' and 'prejudice', which all relate to challenging prejudices around mental illness. Table 7.4 shows examples of these collocates in context.

MYTH:	**People with mental illnesses**	are violent and unpredictable FACT: People with mental illness are more likely to be a victim of violence
Some	**people with mental illness**	say the **discrimination** can be worse than the symptoms
The problem is that if this language is making	**people with mental illness**	feel **stigmatised**, ashamed and isolated then the amount of thought behind it as it is used casually is largely irrelevant.
Mr Miliband will speak out against articles written by Mr Clarkson and broadcaster and journalist Janet Street Porter, which he claims **belittled**	**people with mental illness**	and contributed to a national taboo on the issue.
For many	**people with mental illness**	, stigma is regarded as the single largest obstacle to improving their quality of life

Mrs Brown, from Bath parliamentary Labour party, called for the **prejudice** against	**people with mental illness .**	who wanted to work to be addressed

TABLE 7.4. CONCORDANCE FOR 'PEOPLE WITH MENTAL ILLNESS' WITH CONCORDANCE LINES FEATURING THE STATISTICALLY SIGNIFICANT COLLOCATES '*MYTH*', '*DISCRIMINATION*', '*STIGMATISED*', '*BELITTLE*', '*STIGMA*' AND '*PREJUDICE*'

Table 7.4 shows that the texts in the corpus reveal a high level of awareness of the stigma faced by people with mental illness and even the stigma manifest through language specifically. This awareness is further evidenced by the fact that the lexical item 'stigma' is a relatively frequent word in the corpus, occurring 4,237 times (rel. freq. 72.37 pmw). To contextualise this frequency, 'stigma' occurs just 279 times (rel. freq. 2.48 pmw) in the BNC which contains 10 million words of British English. Furthermore, the frequency of 'stigma' is higher in the MI 1984-2014 corpus than the lemmas 'violent' (rel. freq. 71.14 pmw) and 'criminal' (rel. freq. 52.96 pmw). This is a noteworthy finding when one considers that the majority of previous research into media representations of mental illness have reported that criminality and violence are key themes. This is not to say that violence and criminality are no longer salient aspects of media reports on mental illness, but it does indicate that self-reflective commentary and an awareness of particular linguistic forms and their ideological content is now an established aspect of news reports on mental illness.

The analysis so far has indicated that person-first language correlates with discursive contexts that are supportive of people with mental illness (i.e. contexts in which the article is challenging prejudices and stereotypes about people with mental illness). I will now look at the concordances for the identity-first forms to explore the context in which non-person-first language is used. Table 7.5 shows the top three randomly generated concordances for the identity-first structure 'the mentally ill':

Identity-first examples **'the mentally ill** (Raw freq. 2,071, Rel. freq. 35.37pmw)		
Many of these organizations have been highlighting the stress to both individuals and their families resulting from the early rundown of hospitals for	**the mentally ill**	in advance of adequate provision being made to receive patients into the community.
However, a high-quality health service for	**the mentally ill**	- which would protect patients and society as a whole - requires resources in any event
"We must have equalisation of benefits for	**the mentally ill** ,"	he said. "There is still this terrible slur on the mentally ill

TABLE 7.5. CONCORDANCES FOR 'THE MENTALLY ILL' TAKEN FROM THE MI 1984-2014 CORPUS

Based on the findings of the person-first concordance analysis and the theory behind person-first labels, we might hypothesise that the identity-first forms would occur in contexts that are not as supportive to people with mental illness. However, Table 7.5 shows that this hypothesis does not bear out. Although 'the mentally ill' is listed as a problematic form by anti-stigma initiatives like *Time to Change*, the contexts surrounding this form in the samples taken from MI 1984-2014 are all concerned with bringing attention to adequate medical care and benefits for people with mental illness. At least from these concordances then, identity-first labels do not appear in any stigmatising context as would be expected if, as person-first advocates suggest, identity-first language was *inherently* stigmatising (as proscribing their use suggests). Collocates of 'the mentally ill' are also indicative of this label not being inherently stigmatising. For example, the top five collocates of 'the mentally ill' include 'rights', 'plight' and 'care'. However, if we look in more detail at the collocates of the label 'mentally ill' (e.g. as a premodifer), which is closely related to 'the mentally ill' semantically, we see many more collocates that contribute to stigma such as 'murders', 'homicides' and 'killings'. This finding is indicates that a more stigmatising form than

'the mentally ill' is any naming strategy formed by premodifying a head noun with 'mentally ill'. That is, linguistic evidence suggests that 'the mentally ill', which is a term identified by *Time to Change* as problematic, is arguably less problematic than '[mentally ill + noun]' forms. This is due to the fact that the collocates of the label "mentally ill" reveal that this label is closely associated with criminality. Table 7.6 shows the L1 nominal and adjectival collocates of 'mentally ill' in addition to its R1 nominal collocates.

L1 Collocate (Adj or N)	MI Score	MENTALLY ILL	R1 collocate	MI score
1. Severely	10.62		1. offenders	9.08
2. chronically	9.82		2. inmates	8.20
3. severely	9.42		3. prisoners	8.13
4. seriously	8.50		4. criminals	7.85
5. acutely	8.38		5. offender	7.35
6. elderly	8.28		6. defendants	7.22
7. gravely	8.14		7. persons	7.14
8. seriously	7.90		8. patients	6.96
9. homeless	7.77		9. attacker	6.83
10. elderly	7.33		10. pensioners	6.42

TABLE 7.6. L1 AND R1 COLLOCATES OF 'MENTALLY ILL' (MIN FREQ. 5) IN THE MI 1984-2014 CORPUS

Table 7.6 shows that "mentally ill" often modifies a head noun (shown in the R1 collocate column) that relates to criminality with two of these collocates referring to people who are incarcerated (inmates and prisoners), and who, therefore, have presumably been found guilty of an offence and deemed dangerous to society. This finding is in accordance with previous research into person-first language that found that "the 'modifier + head-N' pattern tends to appear in contexts where the NP refers to 'undesirable' societal elements (e.g., people in prison)" (Halmari, 2011: 838). The fact that 'mentally ill' collocates statistically significantly with lexis related to criminality suggests that criminal offences committed by people with mental illness are over represented in press reports on mental illness (particularly when one

considers that a person with a mental illness is more likely to be a victim of a crime than the perpetrator of one). The ideological effect of this textual association between mental illness and criminality contributes to the view that people with mental illness are dangerous. Furthermore, the L1 collocates of mentally ill such as 'severely', 'seriously' and 'acutely' also arguably present people with mental illness as dangerous because they are not in control of their illness; e.g. they are unpredictable or uncontrollable. Taking the L1 and R1 collocates of 'mentally ill', then, people with mental illness are presented as being out-of-control criminals. An analysis of the identity-first form [mentally ill + noun] indicates that the context is more stigmatising that person-first forms because it misrepresents people with mental illness by overrepresenting criminality in this population. Taken in the context of newspaper discourse, the overrepresentation of articles reporting on offences committed by people with mental illness could be taken as a way for journalists to 'storify' or sensationalise events. On the link between identity-first and news discourse, Halmari (2011: 838) writes "the non-PC syntactic pattern […] seems to be motivated by the editor's desire to make the story more newsworthy".

In the next section, I will discuss the salient naming practices identified in the headline sample.

7.3.2. 'Patients', 'sufferers', and 'victims': Exploring salient naming practices in the headline sample

In section 7.3.1, I showed the distribution of person-first language in the MI 1984-2014 corpus. The reason for this was that there were very few instances of person-first language in the headline analysis. I suggested that the reason for the low frequency of person-first forms in the headline sample was possibly due to the competing aims of journalists, who want to write newspaper headlines which are concise and attention-grabbing, and the aims of person-first language, which is necessarily circumlocutory

(and therefore not well-suited to the conventional structure of headlines). In this section, I will explore the salient naming practices identified in the headline analysis for referring to people with mental illness. I do this to explore what labels are being used in the headlines if person-first forms are not. Specifically, in this section I explore the head nouns 'patients', 'sufferers' and 'victims'[71]. In my analysis, I examine how these head nouns are modified (e.g. "PATIENTS WHO SUFFERED TORMENT"), as well as how these labels pattern across the time period covered in the MI 1984-2014 corpus, and the illness subcorpora.

Of the three labels I discuss in this section, two have been identified by anti-stigma initiatives as problematic. These are 'victim' and 'sufferer'. The reason that *Time to Change* give for deeming these two terms problematic is that they do not accurately reflect the fact that "Many people with mental health problems live full lives and many also recover." (*Time to Change*, 2019). In the headline analysis, 'patient' is the most commonly used term of the three. The label 'patient' is also the most common label of the three in the corpus more generally. Figure 7.3 shows the frequency of 'sufferer', 'victim' and 'patient' across the year subcorpora in the MI 1984-2014 corpus. It is also noteworthy that the frequency of 'victim', 'sufferer' and 'patient' have levelled across the time period, which suggests that the terms are established (i.e. they have started to plateau).

[71] In addition to 'patient', 'victim' and 'sufferer', the label 'the mentally ill' was also a very frequent label in the headline analysis. However, due to the fact that I discuss 'the mentally ill' in Section 7.3.1, I do not discuss it here.

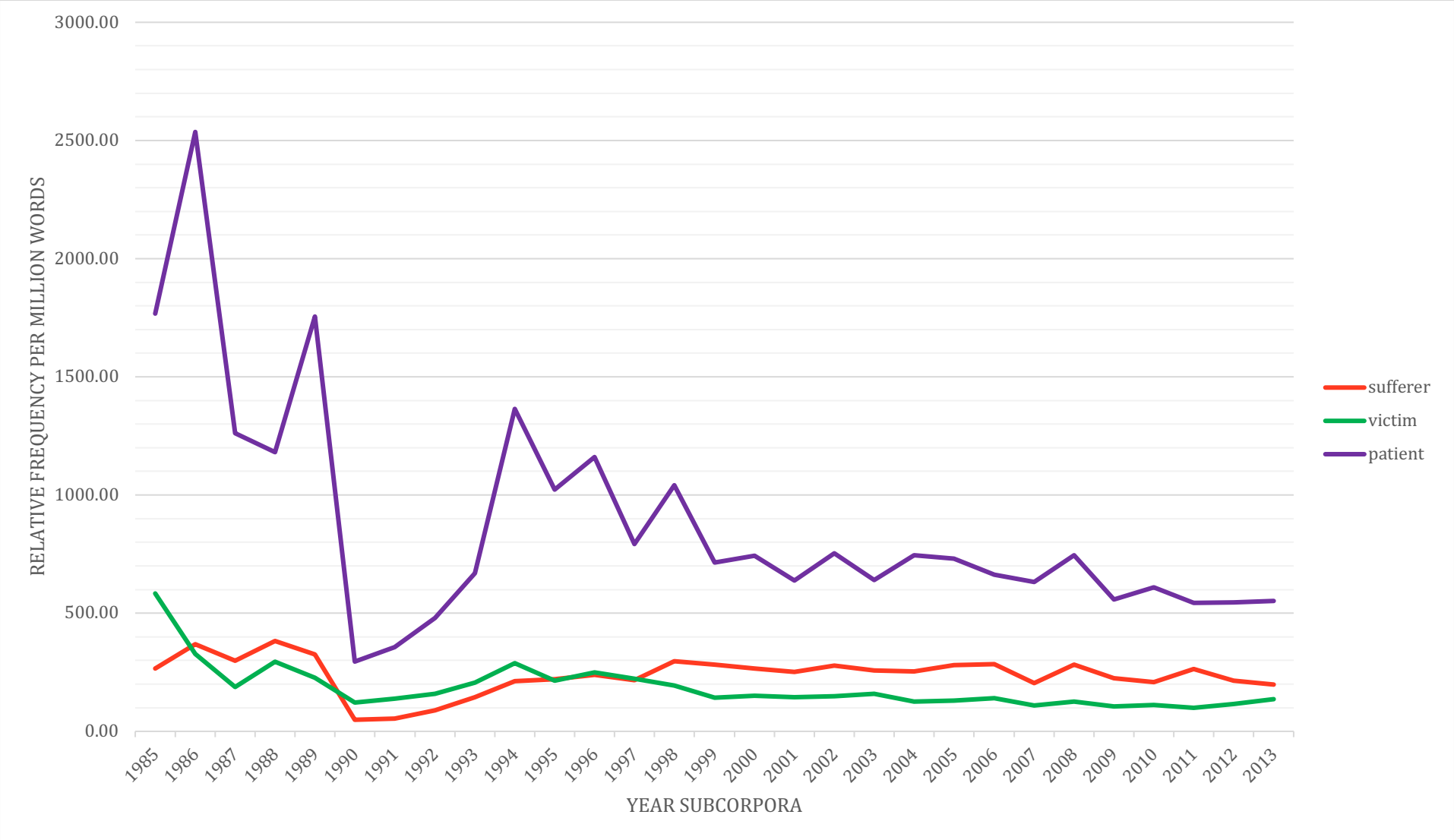


FIGURE 7.4. FREQUENCY OF SUFFERER, VICTIM AND PATIENT ACROSS THE YEAR SUBCORPORA IN THE MI 1984-2014 CORPUS

Figure 7.4 shows that 'patient' is the most frequently used label of the three terms followed by 'sufferer'. Victim is the least frequent of the three. In terms of the semantic content of the three labels, one may intuitively expect this order of frequency (i.e.' patient' being the most frequent, then 'sufferer', then 'victim') because these three terms vary in severity; or, to use Time to Change's words, they vary in the scope the person has for recovery. For example, 'patient' (compared with 'sufferer' and 'victim') is neutral in terms of severity but suggests that the person is undergoing treatment, 'sufferer' is more severe than 'patient' because it suggests the person is actively suffering, and 'victim' is the most severe because it is a label that encodes finality or irreversibility, i.e. once a person is a victim they have no recourse to change their 'victim' status. These three labels, then, may be placed on a cline of severity that corresponds to whether or not recovery is possible. A concordance analysis of the phrase "victim of" in the BNC confirms my interpretation of 'victim' as being the most severe in terms of scope for recovery from an illness. Table 7.7 shows these concordance lines.

1	man who paid the price of gold with a bullet in the back,	**victim of**	Garrimpero gun law
2	TOWARDS the end of her life my mother was in hospital, a	**victim of**	Alzheimer's Disease
3	Rescuers tend a badly-injured	**victim of**	the fireball nightmare
4	THE 3,000th	**victim of**	Ulster's violence paid the price of a brief moment of fame
5	n one recent case the	**victim of**	a stabbing died two and a half years after the incident

TABLE 7.7. CONCORDANCES OF 'VICTIM OF' IN THE BNC

Table 7.7 shows that the context in which someone is described as a 'victim' relates to states that cannot be treated. For example, the thing that a person has been a 'victim

of' is death (specifically murder) in concordance lines one, four and five. In concordance line four, the person is a victim of life-changing injuries, and in concordance line two, the person is a 'victim' of Alzheimer's disease, a disease from which recovery is not possible at this time. Further evidence for the interpretation that 'victim' is often used to refer to people in irreversible circumstances is the fact that, in the sample by year, 'victim' is often used in reference to suicide, e.g. "suicide victims". In addition to 'victim' occurring in contexts where it is not possible to recover from the state or illness described, 'victim' also encodes a sense of powerlessness or passivity on the part of the person described in this way. This is also the case with 'sufferer' because typically people do not suffer from things intentionally (see Chapter 8 for a discussion on the semantics of 'sufferer' and 'suffering'). In order to explore the usage of 'victim' in the corpus, I will now examine in more detail the instances of 'victim' in the headline sample. Table 7.8 shows the instances of 'victim' in the headline sample.

Victim	Victims Victims of Post Traumatic Stress Disorder Victims of winter disorder Survivors who are victims

TABLE 7.8. NAMING PRACTICES IN THE HEADLINE SAMPLE CONTAINING THE HEAD NOUN 'VICTIM'

Of the instances of 'victim' in the headline sample, half relate to PTSD. One of the headlines that relates to PTSD, '*survivors who are victims*', is arguably a reference to the disparity between the sense of 'survivor' and the sense of 'victim', specifically that one person is not typically associated with being both a survivor and a victim. The play on words in this headline constitutes further evidence for the interpretation that 'victim' is less likely to be associated with recovery (e.g. 'surviving' something) than 'patient' or 'sufferers'. A substitution test confirms this interpretation, e.g. 'survivors who are *sufferers*' and 'survivors who are *patients*' are not marked in the way that '*survivors who are victims*' is. This is possibly to do with the association between

victimhood and death e.g. calling a 'sufferer' or a 'patient' a survivor is not marked because these forms are not so closely associated with death in the way that being a 'victim' is. The association of 'victim' with PTSD over other illnesses raises the question of whether there is a patterned usage of particular naming strategies for people who have particular illnesses. In order to explore any such patterns, I plotted the frequency of 'patient', 'sufferer' and 'victim' across the illness subcorpora. Figure 7.5 shows the frequency of each term in each of the illness subcorpora.

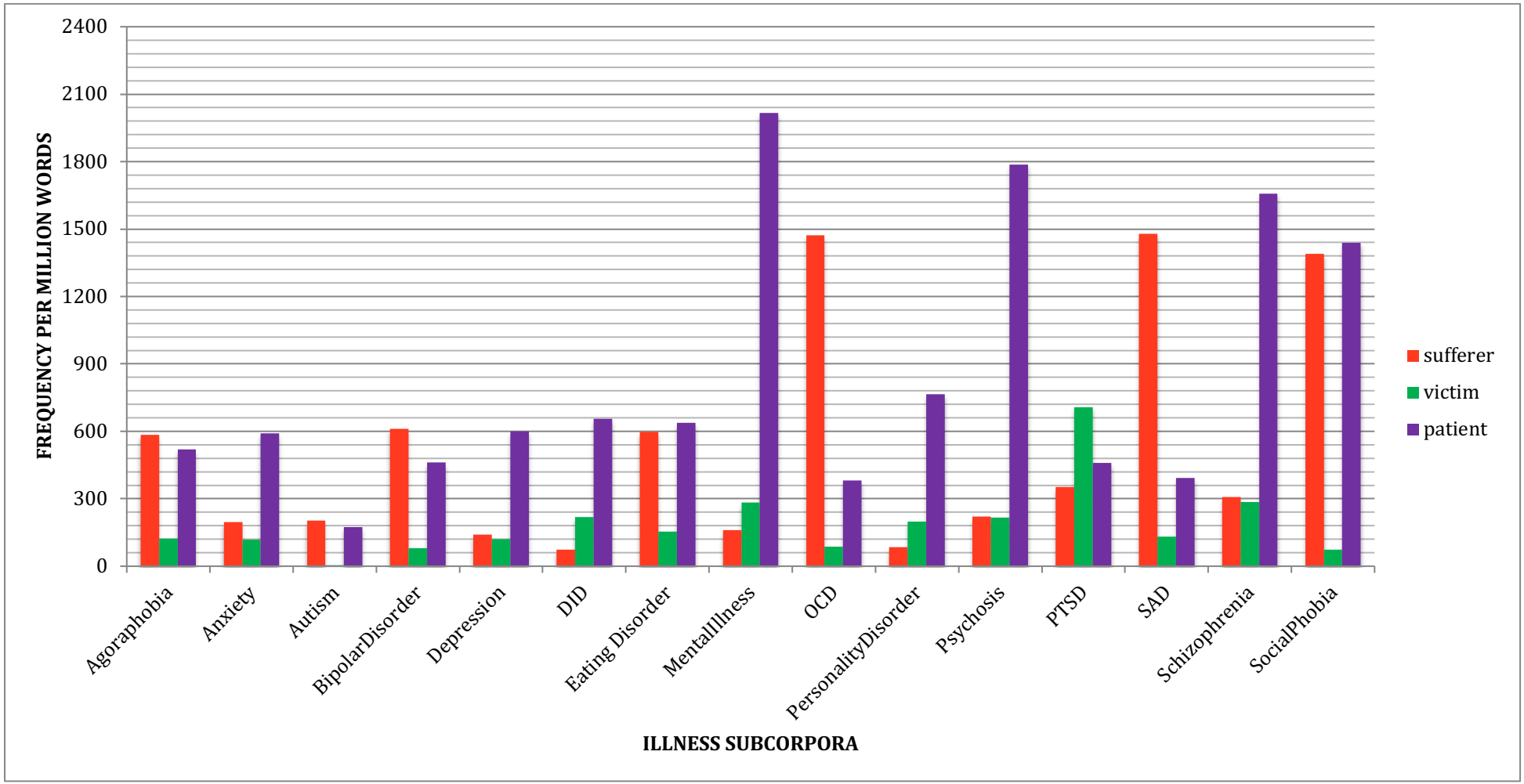


FIGURE 7.5. FREQUENCY OF 'SUFFERER', 'VICTIM' AND 'PATIENT' ACROSS THE ILLNESS SUBCORPORA IN THE MI 1984-2014 CORPUS

Figure. 7.5 reveals that the pattern between describing people with PTSD as victims in the headline analysis is also a pattern in the MI 1984-2014 corpus more generally. In addition to the link between PTSD and 'victim', Figure 7.5 also shows a pattern between 'patient' and the related conditions schizophrenia and psychosis. The symptoms of psychosis include delusions and hallucinations and these symptoms are also the symptoms of schizophrenia. The fact that 'patient' occurs most often in the schizophrenia and psychosis corpora is arguably indicative of these two illnesses being the most pathologised of those represented in the illness subcorpora; i.e. to be a patient denotes medical care, medical care is necessary for the treatment of a condition). The interpretation of 'patient' denoting pathological conditions is supported by collocates of 'patient' in the MI 1984-2014 corpus, which include lexis to do with medical intervention. Collocates of 'patient' include 'doctor', 'professionals', 'GPs' and 'therapist'. A collocation analysis of 'sufferer' and 'victim' reveals that neither term shares these collocates. This provides further evidence that 'patient' denotes pathology in a way that 'sufferer' and 'victim' do not. The pattern between articles on schizophrenia and psychosis and the label 'patient' which suggests that the person is experiencing a pathological illness (or is experiencing "abnormal mental conditions" (OED, 2019)) could be a reason why schizophrenia has been found to be one of the most stigmatised illnesses in the news media (Goulden et al., 2011: 5; Mann & Himlein (2004); Nawka et al., 2012: 1).

Related to the link between 'patients' and medical professionals, an analysis of 'patient' in the headline sample as well as in the Schizophrenia corpus reveals that 'patient' occurs in contexts concerning detention of some sort, particularly against the backdrop of deinstitutionalisation policy in the 1980s, which resulted in a transition from treating people with mental illness as in-patients to 'care in the community', where people were treated in their homes. An example of 'patient' used in the context of community care taken from the headline sample is "second patient free to kill victims" (The Times, Dec 23, 1998). The article reports on a man who killed his neighbour months after being discharged from a psychiatric institution where he was

treated for potential schizophrenia. Collocates of 'patient' in the schizophrenia corpus include the verbs 'discharged' and 'released'. These collocations provide further evidence of a link between the label 'patient' and in-patient care; or to be more precise, 'patients' *leaving* in-patient care to be cared for at home. Sample concordance lines for these collocates (as lemmas) are shown in Table 7.9.

1	But as the movement advanced, more severely ill	**patients**	were **discharged** in need of greater support, and that was lacking
2	**Discharged**	**patients**	who refuse to comply with treatment such as a weekly visit to a clinic for an injection.
3	Health Minister Rosie Winterton denied that the Government was failing to protect people from the dangerous mentally ill, but conceded the Government could not force **discharged** .	**patients**	to take their medication
4	The possibility that the attacker has a history of mental illness again highlighted the grave concern being felt over	**patients**	being **released** under the care in the community programme
5	Many of these highly vulnerable	**patients**	are **released** prematurely, placing them at even more risk of suicide and self harm.

TABLE 7.9. CONCORDANCE LINES FOR 'PATIENT' COLLOCATES 'DISCHARGED' AND 'RELEASED' IN THE SCHIZOPHRENIA CORPUS (L3, R3, MIN FREQ.5)

Table 7.9 shows that the issue of care in the community was a newsworthy item in the corpus overall. The phrase 'care in the community' occurs 992 (16.94 pmw) in the MI 1984-2014 corpus, and 'community care' occurs 1,145 times (19.56) in the MI 1984-2014 corpus. Of the total instances of 'care in the community' in the MI 1984-2014 corpus, 28% (121.95 pmw) occur in the Schizophrenia corpus. Furthermore, 25% (125.91 pmw) of the total instances of 'community care' occur in the Schizophrenia corpus. This

demonstrates that not only is community care deemed newsworthy in mental illness reporting generally, but it is deemed newsworthy in particular reference to schizophrenia.

The concordance lines in Table 7.9 all share similar propositions; specifically they problematise the 'discharge' or 'release' of 'patients' into the community under the care in the community policy. However, only one concordance line (number one), problematises the policy in reference to the support provided by the government (and by extension, local authorities). Concordance line five alludes to the fact that people moving to community care may need additional support (e.g. it refers to patients as 'highly vulnerable' and references the fact that some 'patients' may have experienced discharge from hospital prematurely); however, the focus of the sentence is on the 'vulnerable patients' who are at higher risk of 'suicide and self-harm' rather than the reasons why these people were prematurely discharged from hospital without adequate home support. The rest of the concordance lines do focus on the discursively constructed 'patient' however. Concordance lines two and three discuss patients (consciously) not taking their medication, e.g. the complex noun phrase (comprising an NP and a PP) that constitutes concordance line two "**Discharged patients** who refuse to comply with treatment such as a weekly visit to a clinic for an injection." The reporting of patients discharged from hospital not taking their medication is represented in the verb phrases encoded in one and two as being something that 'patients' are wilfully not doing:

Concordance line two:	**Discharged patients** who *refuse* to comply with treatment
Concordance line three:	the Government could not *force* **discharged patients** to take their medication

As indicated above (see italics), the head verbs that the journalists have chosen both encode a conscious choice on behalf of the people with mental illness to not take

medication (when, in fact, it could be the case that people are not taking medication due to failures in the community care system rather than from a desire to not engage with treatment (Rose, 1998)). Further, concordance lines two and three feature logical presuppositions in that they presuppose that (a) discharged patients should comply with treatment, and (b) the government should be able to force people to take medication. Taken together, the presentation of people with mental illness (here 'patients') as being out of control, and the implication that the government should be *in* control, contributes to the idea that people with mental illness are dangerous people in the community rather than active agents who can negotiate their own care in partnership with care workers if necessary. The link between presenting people with mental illness as violent and the community care policy contributes to the attitude that community care "has resulted in violent madmen being let loose to roam the streets and prey on an innocent public" (Rose, 1998: 213). The presentation of people with mental illness living in the community as a threat to the public can be seen in concordance line four, below (this theme is also present in concordance three where people with mental illness are referred to as "the *dangerous* mentally ill"):

Concordance line four:	The possibility that the attacker has a history of mental illness again highlighted the grave concern being felt over **patients** being **released** under the care in the community programme

In concordance line four, patients living in the community under the community care policy are being cited in reference to a violent attack committed by someone who may or may not have a mental illness. Specifically, the article discusses the possibility that the care in the community policy will increase future attacks by mentioning previous events in which violent offences have been committed by people with mental illness. Concordance line four, then, includes the speculation that community care will increase violent attacks (using historical and isolated events) in addition to speculating about whether the attack being reported was actually committed by a

person with mental illness (indicated by 'the possibility that the attacker has a history of mental illness). As a result, the article links violence and mental illness, as well as an increase in violence and the community care policy, based entirely on speculation. Given that the news media constitutes a significant source of public information on mental illness, such speculation perpetuates damaging and inaccurate ideologies about people with mental illness on a large scale, and as a result contributes to stigmatising attitudes about people with mental illness. To return to the notion of storification of events in news articles, it is hard to view the kind of speculation shown in concordance line four as serving any other purpose than to sensationalise the event in the absence of newsworthy facts. Such sensationalist reporting may not just result in stigmatising attitudes about people with mental illness however. Previous research has shown that sensationalist reporting can create societal panic. This societal panic can then influence mental health care policy (Paterson, 2006). Moreover, such sensationalist reporting may hinder community integration because, as Hannigan (1999: 431) states, "Public tolerance of, and non-discrimination towards, people with mental health problems are key factors on which success in achieving the goal of community-based mental health care".

In this section I have described the salient naming practices associated with mental illness reporting. I have discussed the frequency of these labels and explored how they are patterned across the time period and the illness subcorpora. In the next section I explore themes in the naming practices identified in the headline sample and the sample by year. I enrich my analysis with examples taken from the MI 1984-2014 corpus.

7.3.3. 'Groups', 'cases' and 'the 1 in 4': referring to people with mental illness as quantities

In this section I describe a salient theme identified in the data analysis. This theme is

the tendency in the corpus to refer to people with mental illness as quantities; for example, 'one in four adults' or 'x per cent of the population'. Although my interest is still related to the linguistic analysis of naming strategies used in the press reports on mental illness, in this section I take a wider view of naming to include the ways in which people with mental illness are referred to.

I have previously stated at several points in this chapter that there is a tendency in the corpus to refer to people with mental illness in the plural. For example, I showed in Section 7.3.1 that '*people* with mental illness' was more common that '*person* with mental illness'. Furthermore, the randomly sampled concordances shown for the lemma 'patient' in Section 7.3.2 were all in the plural form. So too, were all the instances of 'victim' shown in Table 7.8. It is the case, in fact, that all three of the labels explored in Section 7.3.2 ('victim', 'patient' and 'sufferer') are more commonly used in the plural form in the MI 1984-2014 corpus. Table 7.10 shows the frequency of the three labels in their plural and singular form.

Label	Freq. in singular	Freq. in plural	% of total (plural)
Patient	8,916	23,386	72%
Victim	3,531	4,496	56%
sufferer	2,227	10,688	82%

TABLE 7.10. FREQUENCY OF SALIENT NAMING STRATEGIES USED IN PLURAL AND SINGULAR FORM IN THE MI 184-2014 CORPUS.

Table 7. 10 provides further evidence for the fact that there is a very strong tendency in media reports on mental illness to refer to people with mental illness in the plural. In addition, there is also the related tendency in the corpus to refer to people with mental illness as quantities or statistics. The headline analysis provides examples of this tendency:

People as numbers/ statistics	Numbers One in four One in four adults One in four of us HALF US POPULATION

TABLE 7.11. REFERRING TO PEOPLE AS NUMBERS OR STATISTICS IN HEADLINE SAMPLE

The reference to the statistic 'one in four' relates to how many people will experience a mental health problem in their life, and is frequent in the headline sample and relatively frequent in the corpus more generally (1595 instances, 27.24 pmw). In addition to being a prevalent feature of the headline analysis, I identified the practice of referring to people with mental illness as quantities or statistics as a salient feature of the sample by year. Table 7.12 shows some examples of these prevalent themes. All the examples were used in context to refer to people with mental illness:

People as quantities or statistics in the sample by year	
Numbers/Statistics	A worrying **number** of people **more than three quarters** **one third** of the population **1.5m** Brits with an eating disorder **36,729** people **21,058** who have mixed anxiety or depression **8,955** with a general anxiety disorder **19** suicides **an increasing number** of men **two thirds** of all female prisoners **1 IN 3** YOUNG WOMEN **one in three** 16- to 25-year-old **one in four** people **one in five** older people **one in six** **1 in 200** women affected by anorexia women **twice as likely** to suffer from depression **40 per cent** of young **10%** of all women **5%** of men

	54 per cent of women aged 16 to 25
Quantities	The **majority** of the mentally ill **some** patients **most** schizophrenic patients
Nouns denoting groups	this **group** of people The **remainder** the **group** of patients **cases** of mental illness more **cases**

TABLE 7.11. PEOPLE DESCRIBED AS QUANTITIES OR STATISTICS IN THE SAMPLE BY YEAR

As Table 7.12 shows, there are many examples in the sample by year that attest the finding that people with mental illness are often referred to as quantities or statistics. In addition to being referred to as quantities in terms of numbers or percentages (e.g. one in six, 5% of men), people with mental illness are also referred to in this way via nouns or determiners denoting quantity (e.g. 'the majority of', 'some patients'). Furthermore, people are also quantified through the use of nouns denoting groupings or quantities (e.g. 'cases', 'this group'). Taken together, Tables 7.10 and Table 7.11 show that people are discursively grouped or quantified through morphological marking (i.e. plural suffixes), as well as at the lexical level (e.g. through lexical items

denoting groups). Further evidence for this tendency is the fact that an n-gram analysis of the top 100 5-grams in the MI 1984-2014 corpus show that percentages, numbers and describing the different states between numbers (e.g. increase, rise) is a salient feature of the corpus.

Rank	5-gram	Raw frequency
15	Per cent of the population	665
30	Million people in the UK	448
34	Increase in the number of	423
65	Rise in the number of	310

TABLE 7.12. 5-GRAMS RELATING QUANTITY OR NUMBER IN THE MI-1984-2014 CORPUS

It could be proposed at this stage that the tendency to report issues in the news in terms of percentages and quantities is a feature of newspaper discourse more generally because, as I have previously stated, part of the 'storification' of events involves establishing the event in some social context (for example the speculative links created between violence and schizophrenia described in Section 7.3.2). However, an analysis of the top 100 5-grams in the SiBol Corpus (a corpus of 650 million words comprising 1.5 million articles from UK broadsheet newspapers between 1993-2013) reveals just one shared 5-gram ('increase in the number of') which occupies a fairly low rank (no. 72) in comparison to its rank in the MI 1984-2014 corpus (no. 34). This suggests, then, that the quantification of people (and things more generally) is not just a feature of newspaper discourse, but specifically a feature of the newspaper discourse contained in the corpus. Because of this, we may hypothesise that quantification (e.g. the overrepresentation of statistics) is a feature of mental illness discourse.

The n-gram analysis is also revealing of another feature of quantifying people and entitles related to mental illness – that of shifting quantities, particularly verbs denoting increasing numbers such as *increase* and *rise*. A concordance analysis of the collocates that modify 'rise' reveal that the nature of the increase reported in the

corpus is in the majority of cases a sharp increase. Table 7.14 shows the modifiers of 'rise' that denote a sharp increase in the MI 1984-2014 corpus:

Modifier	MI score
sharply	10.83
dramatically	10.01
rapidly	9.36
steeply	8.79
tenfold	8.46
fast	8.33
inexorably	8.09
significantly	7.88
threefold	7.83
exponentially	7.81
substantially	7.75

TABLE 7.13. COLLOCATE MODIFIERS OF 'RISE' IN THE MI 1984-2014 CORPUS

Table 7.13 shows that not only are numbers and statistics a prevalent feature of news articles on mental illness, but so too is the reporting of those numbers as increasing in unprecedented ways. A closer inspection of the modifiers of 'rise' in context reveals that the things 'rising' include (but are not limited to) mental illness, psychotic problems after drug abuse, autism, suicide and self harm, and concern. Further examination of the verbs that collocate with 'mental illness' shows that both 'rise' and 'increase' are statistically significant collocates. Table 7.14 shows some example concordance lines featuring 'mental illness' + 'rise'.

He repeated the exercise in 1985 and found the number with 'psychiatric morbidity' which means they showed symptoms of	**mental illness**	had **risen** from 22 per cent to nearly a third
the number of murders committed by people with	**mental illness**	**rose** from 54 in 1997 to over 70 in 2004 and 2005
nationally the prevalence of	**mental illness**	has **increased** by 14%

It is not surprising that officers are dealing with more incidents involving mental health issues as the incidence of	**mental illness**	is **increasing** and more people are being cared for in the community.

TABLE 7.14. EXAMPLE CONCORDANCE LINES OF [MENTAL ILLNESS + RISE/INCREASE] IN THE MI 1984-2014 CORPUS

Table 7.14 shows that it is not just the case that there is a tendency in the corpus to refer to people in quantities, but that there is also a tendency to quantify the incidence of mental illness. Specifically, there is a tendency to report that the rates of mental illness are rising (and therefore the number of people with mental illness is rising). As shown in Table 7.14, the nature of this rise is presented as being fast, significant in number and (as indicated by 'inexorably') in a manner that is out of control or impossible to prevent. This is despite the fact that the "overall number of people with mental health problems has not changed significantly in recent years" (Mind, 2019a). Taken in the context of violence and criminality that I explored in the previous section, what the over-representation of rising quantities results in is the implication that the rise in cases of mental illness will result in a rise of violence or criminality. Furthermore, this reported rise is one that is purportedly out of control. The link between a rising incidence of mental illness and criminality and violence, then, looks on the surface to be a logical extrapolation from facts; however, these links are based on untruths because there is no evidence that the incidence of mental illness has increased significantly (at least since in recent years as Mind (2019a) state. Furthermore, the rise in numbers, statistics and quantities reported are never contextualised; e.g. statistics about the incidence of mental illness are never reported alongside statistics about other types of illness, and statistics concerning violent crime committed by people with mental illness are never reported alongside statistics on violent crime committed by people who do *not* have a mental illness. As a result, it is impossible for the average reader to comprehend these rising numbers in context. To

return again to the notion of 'storification' that I presented in Section 7.1, I argue that the overrepresentation of decontextualized, rising quantities by journalists sensationalises or 'storifies' events. Specifically, the overrepresentation of numbers and statistics provide a means for journalists to create a moral panic (Cohen, 1972) around mental illness. I will return to the sociological notion of moral panic later in this section.

The overrepresentation of statistics and numbers (and in particular rising numbers) provides a basis for journalists to speculate and create links between events and mental illness that may not exist. Recall, for example, the way that the journalist speculated about whether an attacker was mentally ill and used that conjecture to further speculate about increase risk of violence committed by people with mental illness more generally:

> The possibility that the attacker has a history of mental illness again highlighted the grave concern being felt over **patients** being **released** under the care in the community programme

As a result of the overrepresentation of increasing quantities, speculation like that shown in the concordance line above is harder to spot (and therefore question) because there is a societal assumption (created through language) that people with mental illness pose a threat to society. This threat is not just the result of the overrepresentation of increasing quantities, but is the composite effect of this overrepresentation and the problematic naming strategies identified in the previous sections (e.g. the link between being 'patient' and criminality). The type of speculation shown in the concordance line above, then, is supported by, and provides support for, the false belief that the number of people with mental illness is rising and that this rise poses a risk to society.

In this section, I have systematically analysed the occurrence of numbers and statistics in the corpus. I did this in an exploratory fashion that was not guided by any particular theoretical angle. As a result of the findings of this exploratory analysis, I raised the hypothesis that the overrepresentation of people as quantities was a feature of news reports on mental illness because such quantification was not a feature of general newspaper corpora. However, previous linguistic research has shown that the "rhetoric of quantification" (Fowler, 1991: 166) is not necessarily a feature of news reports on mental illness, but one way that journalists contribute to "press hysteria" (Fowler, 1991: 146). In fact, all of the features I have described up to this point are in line with Fowler's (1991) analysis of "press hysteria". In an analysis of the news reportage on the salmonella outbreak[72], Fowler (1991) identified "the rhetoric of quantification" as the "dominant stylistic feature" of "press hysteria" (1991: 165). One such linguistic feature of 'the rhetoric of quantification' identified by Fowler (1991: 168) was the prevalence of "verbs, or nouns derived from verbs – designating changes in numbers". We can now see that the salient 5-grams that denoted number changes described above are also a contributing factor in the creation of press hysteria surrounding mental illness. Fowler (1991) writes of the rhetoric of quantification that

> The result is a blurring, a diminution in analytic precisions; an impressionist style comes over, especially in conjunction with the ubiquitous mentioned of large but constantly shifting numbers. The discourse is constantly alarming and hyperbolic but in an obscure way: a problem of considerable propositions is always being alleged; we are bound to be concerned about it, but its outlines are indistinct, like some huge threatening shape on the horizon in a bad horror movie
>
> (Fowler 1991: 168-169)

[72] The salmonella outbreak refers to a public scare in the UK in 1988 caused by Edwina Currie MP who was then a Junior Health Minister. Currie quoted some unpublished government statistics on national television about the number of eggs that contained the salmonella bacteria which can cause serious food poisoning.

Fowler's (1991) findings indicate that the rhetoric of quantification is not simply a feature of mental illness then, but rather a feature of news reports on issues that are deemed to be a threat to society. Fowler's (1991) analysis, however, is concerned with the threat that bacteria poses, and not people. The rhetoric of quantification has much wider and more serious implications when used in reference to mental illness because this "huge threatening shape" with "indistinct outlines" described by Fowler *does* take shape when applied to the language used to report on mental illness – it takes the shape of people, with thoughts and feelings, who are also consuming these logical falsehoods and, as I reported in the introduction, are internalising these negative representations and feeling unable to seek help as a result.

I have shown in this section so far that the grouping or quantification of people with mental illness is linguistically encoded at the morphological level (e.g. the overuse of plural forms on nouns describing people with mental illness) as well as at the lexical level (through the overrepresentation of numbers, statistics, and words denoting groups (e.g. cases, groups). In addition to these linguistic features of the 'rhetoric of quantification' there are also specific naming strategies that imply that the rate of mental illness is increasing. Take for example the following way of naming mental illness in the headline sample and sample by year:

(1) THE SEROTONIN SOCIETY (1997)

(2) The Prozac Generation (2008)

These two headlines also have the effect of presenting the incidence of mental illness as much higher than it actually is. Specifically, in (1) the implication is that so many people are depressed that we have become a society that needs serotonin (serotonin is the chemical in the brain that contributes to feelings of happiness). The idea of mental illness being widespread in society is also the implication that underlies (2) which is the headline of an article reporting on the number of people taking the anti-depressant Prozac. The headline reiterates the message of the article which is that anti-

depressant use is so widespread that it has resulted in a whole generation of people that take Prozac. Both the headlines are existential presuppositions, i.e. they presuppose the existence of the entities that they discuss, which has the effect of representing them as realities. Existential presuppositions are typically comprised of the definite article + NP, as is the case in (1) and (2). The use of the definite article in (1) and (2) presents 'the serotonin society' and 'the Prozac generation' as given information. As a result, these two entities (the serotonin society and the Prozac generation') are discursively constructed as things that are recognisable to the assumed readers because they are "presupposed to exist by virtue of being in definite noun phrases" (Jeffries, 2010: 95). Compare, for example, how the meaning of (1) and (2) are permuted if the NPs are broken down and the definite article swapped for the indefinite article:

The serotonin society → a society that is dependent on serotonin
The Prozac generation →a generation that is dependent on Prozac

By changing the structure of (1) and (2) the implications they contain (that the whole of society is depressed, and the whole of a generation is depressed, respectively) is easier to question because what was presented as given information (contained in a single NP) is now presented as new information (in a relative clause). As a result, the new information is syntactically constructed as circumstantial information about the entity the noun describes (society, a generation) rather than as information that comprises the entity described by the noun. The headlines constitute another example of how the prevalence (existing cases) and incidence (new cases) of mental illness (in this case depression) are overrepresented. An extract from the 'Prozac Generation' article demonstrates that the theme of quantification is also present in the main body of the article:

> In 2006, the NHS issued **31 million scripts** for Prozac in the UK. Anti-depressant prescriptions cost the health service **£3.3 billion** last year, **almost three per cent** of the entire NHS budget. But research has surfaced that claims they have little effect on mild depression and we may as well have lifted the mood making paper chains with all that cash.
>
> (*Daily Record*, February 29, 2008)

The fact that the main body of the article contains statistics that the headline relies on provides further evidence that the rhetoric of quantification is not just a feature of headlines but it also forms part of what makes articles coherent. The extract also provides another example where it becomes harder to question one part of the article (here the headline) because of the sheer number of statistics (and numbers relating to all different entities) contained in the main body of the article. Furthermore, the extract above provides another example of how numbers and statistics related to mental illness are not contextualised in news articles (i.e. the amount of scripts and money spent on other illnesses are not included in the article to allow for comparison). Moreover, the presentation of these facts as newsworthy suggests that the expense incurred in treating mental illness is not justified, which raises a question as to why this would be the case. Taken in the context of the findings reported in this chapter so far, then, press reports create a paradoxical situation in how they report on mental illness: the number of people with mental illness is rising (which poses a threat to society) and yet the treatment of mental illness is not a worthy way to spend NHS money.

Earlier in this section I referred to the tendency for texts in the corpus to label people and entities related to mental illness as quantities as contributing to a 'moral panic' about mental illness. In light of the analysis conducted so far in this section, I will now conclude this section by arguing that the *linguistic* analysis of news reports I have conducted in this chapter provides further evidence to support the findings of research conducted outside of linguistics that news reports on mental illness

constitute a 'moral panic' (Pearson, 2000). Cohen (1973) defines a period of moral panic as

> A condition, episode, person or group of persons emerges to become defined as a threat to societal values and interests; its nature is presented in a stylized and stereotypical fashion by the mass media
>
> (Cohen 1973: 2)

Cohen identifies three ways that events or issues are reacted to in the media that contribute to the creation of a moral panic. These strategies are: "exaggeration and distortion", "prediction" and "symbolization". (Cohen, 1973: 25). The first strategy, 'exaggeration and distortion', relates to "exaggerating grossly the seriousness of the events […] such as the number taking part, the number involved […] and the amount and any effects of any damage or violence" (Cohen, 1973: 26). As I have shown at various points in this section, this strategy is visible in the MI 1984-2014 corpus through the tendency to overrepresent the number of people with mental illness, including numbers related to incidence, numbers related to those seeking treatment, and numbers related to crimes committed by (or even *possibly* by) people with mental illness. Moreover, figures reported in news articles on mental illness are distorted because they rarely include any comparable figures, which means that the figures are not clear enough to comprehend in any meaningful way for the average, non-specialist reader. In addition to the features of the corpus that contribute to the 'exaggeration and distortion' of mental illness, the second strategy for creating moral panic, 'prediction', is also present in the corpus. Prediction refers to "the implicit assumption […] that what happened was inevitably going to happen again" (Cohen, 1973: 35). Due to the fact that there is an overrepresentation of numbers with no comparative figures provided, predictions can be made that are not easily questionable because the 'exaggeration and distortion' stage has resulted in mental

illness being described as an existential problem. The example I gave earlier in which a journalist speculated about supposed future attacks committed by people with mental illness constituted a prediction. In the example I gave earlier, the journalist speculated that an attack may have been committed by a person with mental illness (whether the person did have a mental illness was not known at that time). The journalist then used that speculation to state that there was "grave concern" in society that there would be more violent attacks committed by people with mental illness. Finally, 'symbolization' refers to the process by which "neutral words […] can be made to symbolize complex ideas and emotions" (Cohen, 1973: 36). Cohen (1973) discusses symbolization in relation to concrete nouns such as place names (which become associated with particular events), but I argue that salient naming practices such as those I identified in Section 7.3.2, for example 'patient', also constitute symbolization. The reason for this is, as I showed in Section 7.3.2, the use of particular naming strategies is patterned and the use of particular strategies can therefore symbolize something much more complex than their surface form would suggest, e.g. the link between the label 'patient' and schizophrenia.

In addition to the symbolization of labels such as 'patient', symbolization also occurs in the corpus in relation to the names of people with mental illness who have committed violent crimes, or where the names of the individuals symbolize a social issue. A salient example of this is the label 'Clunis' (which I briefly mentioned in Chapter 2 and which has attracted much media and academic attention; see Hallam, 2002 and Rose, 1998). 'Clunis' relates to the case of Christopher Clunis, who was diagnosed with schizophrenia and who killed another man in London. The 'Clunis case' attracted much media attention and was framed as an example of the failures of community care. As a result, 'Clunis' is used in the corpus to symbolize the "complex ideas and emotions" (Cohen, 1973: 36) associated with community care and violence committed by people with mental illness. In addition to the Clunis case, another oft-cited case used by the press to discuss the failures of community care is that of Ben Silcock. Silcock, who had schizophrenia, climbed into the lions' enclosure at London

Zoo whilst experiencing a schizophrenic episode and was seriously mauled. Table 7.14 shows examples of the symbolization of 'Clunis' and 'Silcock' to refer to failures of community care (both 'Clunis' and 'Silcock') and to violence committed by people with mental illness ('Clunis' only):

Why is it we only ever hear about the mad axeman; the	Christopher **Clunises**	and those who drive the wrong way up a motorway inflicting untold damage on anyone who happens to get in their way?
It was clear that the system was failing the	**Clunises**	of this world almost as brutally as it had failed Jon and me
There are hundreds of thousands - yes, hundreds of thousands - of	Ben **Silcocks**	in our land
We can and must do more for the quarter of a million	Ben **Silcocks**	of Britain and for their families.
Down almost any street in this land there are	Ben **Silcocks**	

TABLE 7.15. EXAMPLES OF 'CLUNIS' AND 'SILCOCK' USED TO SYMBOLIZE VIOLENCE COMMITTED BY PEOPLE WITH MENTAL ILLNESS

In this section I have shown that people and entities related to mental illness are routinely quantified and presented as statistics. I have showed how the overrepresentation of statistics and numbers contributes to 'press hysteria' (Fowler, 1991) and creates a 'moral panic' around mental illness. However, an ideological effect of the overrepresentation of quantities in the corpus is not just that such reporting presents mental illness as a bigger issue or a bigger threat to society than it is, but also that the quantification of people necessarily detracts from the presentation of people as individuals. I argue, then, that in addition to contributing to 'press hysteria', the

'rhetoric of quantification' that Fowler (1991) identifies (which I argue is present in the naming strategies I have identified) also constitutes a 'rhetoric of depersonalisation'. What I mean by this is that the systematic representation of people as numbers backgrounds the individual experiences of people. As a result, it is much less likely that readers will empathise with individual people and their experiences of mental illness.

In the next section I offer some final discussion and conclusions arising from the analysis I have conducted in this chapter.

7.4. Conclusion

In this chapter I have argued for the particular value of naming analysis in mental illness discourse. I have shown that, in line with previous research, naming practices constitute a linguistic basis for discrimination and therefore can perpetuate stigma about certain populations. I explored the naming strategies that arose from the qualitative analysis of the headline sample and sample by year, in addition to the frequency of naming strategies prescribed by anti-stigma initiatives such as *Time to Change*. In my analysis I showed that overall the use of person-first language is less frequent than non-person-first language. However, corpus evidence shows that person-first forms are rising. This rise suggests that prescribed forms are being more widely adopted by journalists writing about mental illness. In addition, I showed that person-first forms occur in contexts that are supportive of people with mental illness, which I argue constitutes a relationship of correlation between person-first language and positive representations of mental illness (rather than causation, as person-first advocates suggest). The reason for this was that positive representations of mental illness were also found to occur in with non-person first labels. The specific example I gave to exemplify that person-first labels were not the only naming strategy to occur in positive contexts was the label 'the mentally ill', which *Time to Change* advise

journalists to avoid on the basis that it is a non-person first and therefore stigmatising form. I showed in Section 7.3.1 that in comparison to 'the mentally ill', the adjectival modifier ''mentally ill' is a label better to avoid because it is typically followed by nouns that link mental illness, violence and criminality.

In addition to the findings I reported in reference to person-first language, I also showed that the labels 'patient', 'victim' and 'sufferer' are salient head nouns in naming practices for people with mental illness. I showed that there are semantic differences in the naming strategies that contribute to the frequency of each term; e.g. 'patient' is pathologized and 'victim' has connotations of irreversibility. Related to the sematic differences between each word, I also showed that the use of particular naming strategies is patterned across the mental illness subcorpora. Specifically I showed that if we are to base the naming practices for each illness on the frequency of each term in the relevant illness subcorpus, people with PTSD in the corpus were most likely to be labelled as 'victims' and people with schizophrenia were most likely to be labelled 'patients'. I also discussed how the semantic associations 'victim', 'patient' and 'sufferer' carry may contribute to how people with certain illness are perceived. For example, schizophrenia 'patients' are constructed as dangerous to the community.

In the final section of the analysis in this chapter, I showed that people and entities related to mental illness are routinely labelled as statistics or quantities which are presented as rising both quickly and significantly. I argued that this overrepresentation of statistics constituted what Fowler (1991) calls 'the rhetoric of quantification' which contributes to a 'moral panic' (Cohen, 1973) around mental illness. I argued that the quantification of people depersonalizes them and that for this reason, the overrepresentation of people as numbers was not just a feature of 'the rhetoric of quantification' but also is a feature of the rhetoric of depersonalisation.

In the next chapter I explore the ways that 'having' mental illness is encoded in language, particularly in reference to Hallidayean transitivity analysis. In Chapter 7 I also revisit the notion of prescribed linguistic forms but in reference to processes (i.e. verbs) rather than naming practices.

8. 'Suffering' illnesses and 'experiencing' symptoms: ways of talking about having mental illness

8.1. Introduction

In this chapter I address research questions 3 and 3.1. These are "What processes are associated with mental illness?" and "What terms do the press use to refer to having mental illness". Specifically, I explore the ways in which the press talk about 'suffering' from mental illness and 'experiencing' mental illness. The reason for focussing specifically on these two processes only is due to the fact that I identified 'suffering' as an interpretatively significant and frequent collocate in the initial qualitative analysis (detailed in section 8.4), which warranted further lexicogrammatical analysis. The reason for exploring the verb 'experiencing' is that anti-stigma initiatives have identified 'experiencing' as the preferred term for journalists to use when writing about people having mental illness (*Time to Change*, 2019).

As stated in Chapter 4, transitivity analysis is concerned with how authors encode meaning in texts, specifically how they choose to represent actions, events and states. Transitivity analysis is an analysis of the "clause as process" (Halliday, 2003: 315), and unlike a purely syntactic grammatical analysis, which is concerned with "position and sequence of elements, rather than their propositional meanings and functions" (Fowler, 1991: 77), transitivity analysis allows for the analysis of the 'semantic configurations' of structures of meaning (Fowler, 1991: 71). As a result, transitivity analysis allows for the systematic analysis of ideology in texts because it allows the researcher to question why particular linguistic choices were chosen over others. For example, it can offer insights into why a journalist may choose to represent the state of having mental illness as 'suffering' from it rather than 'experiencing' it. As

Fowler (1991) states, "transitivity has the facility to analyse the same event in different ways, a facility which is of course of great interest in newspaper analysis" (Fowler, 1991: 71). The possible ways that a journalist may choose to represent actions, events and states are of particular interest in the analysis of newspaper discourse. This is because, as Hall (1978) writes, "the media do not simply and transparently report events which are 'naturally' newsworthy *in themselves.* 'News' is the end-product of a complex process which begins with a systematic sorting and selecting of events and topics according to a socially constructed set of categories" (Hall, 1978: 53). Adopting transitivity analysis for the purposes of analysing what processes and participants are associated with mental illness in the press, then, can also be said to give an indication of the way that those participants and processes are constructed in social life.

Section 8.2 will outline the method for the data collection and analysis in this chapter. In section 8.3, I discuss the experiential basis of language in reference to the language prescribed by anti-stigma initiatives. In section 8.4, I analyse 'suffer' in the most prototypical texts for each year and explore the verb 'experience' as an alternative to 'suffer'. In section 8.5, I explore the lexicogrammatical features of 'suffer', 'experience' and 'suffer from' to demonstrate that each of these usages functions differently in the MI 1984-2014 corpus and in general language corpora. In section 8.6, I discuss my findings and conclude.

8.2. Method for this chapter

Due to the fact that transitivity analysis sits between a semantic and a syntactic description of a language, it is not easily automated because it does not always deal with elements of linguistic form. The decision about whether a particular verb constitutes one process type or another requires a detailed exploration of the surrounding context of the verb in that particular context. For this reason, the analysis in this chapter will be much more qualitative (and therefore smaller in scale) than the

analyses conducted in the other analysis chapters. This zeroing-in on the data is necessary for transitivity analysis but is also beneficial to the project as a whole due to the fact that both micro and macro analysis is essential for the analyst to build an accurate picture of what phenomena are contained in the data under investigation. Moreover, as I have stated at various points in this thesis, one of the innovations of my work here is that the analysis is integrated. That is to say that it takes account of both the micro (for example the use of a particular verb in a particular instance) and macro (for example keywords in the corpus compared with a reference corpus) features of the data. Practically I have combined qualitative and quantitative methods by using the corpus to identify areas of the data for qualitative analysis; I have also used qualitative analysis in order to inform quantitative analysis, i.e.

Quantitative to qualitative	→	e.g. exploring statistically significant lexical items through concordance analysis
Qualitative to quantitative	→	e.g. noticing an interpretatively significant usage in an article and exploring its distribution in the corpus as a whole

In order to downsize the corpus to conduct the transitivity analysis, I used stratified sampling by year, using ProtAnt (Antony & Baker, 2017) to identify the most prototypical text for each year included in the MI 1984-2014 corpus. ProtAnt calculates the most prototypical text (i.e. individual newspaper article, in my case) in a corpus based on the number of keywords the text contains when compared to a reference corpus (comprising all other years). This means that the text that contains the most keywords will be the most prototypical for that year.

Once I had collected the most prototypical texts (n=31), I then selected sections of the articles for further analysis guided by how relevant the part of the text was to exploring the research question (i.e. What processes and participants are associated with mental illness?). I then analysed each extract for interpretatively significant processes. This process led to the identification of 'suffer' as an interpretatively

significant collocate in the sample (occurring in just over a third of all the articles). The next section describes the theoretical basis for the analysis I conduct in this chapter.

8.3. The experiential basis of language and the linguistic basis of experience

In the introduction to this chapter, I described transitivity as an analysis of the way actions, events and states are represented in language. What I did not explicitly state, which is vital to my point in the analysis here, is that from exploring those actions/events and states we obtain insight into human experience; that is, how people perceive the world and their experiences within it. In his 1998 article, 'On the grammar of pain', Halliday explores the various ways that pain can be created in language, specifically pain as a participant and pain as a process. On the utility of transitivity analysis for exploring words such as 'pain', he writes:

> The grammar of every natural language is a theory of human experience, and it is a powerful theory in that it covers every aspect of that experience both real and imaginary; yet pain does not fit easily and naturally into the phenomenological model the grammar provides, despite the fact that it has obviously been a part of it from the beginning. But, on the other hand, and for that very reason, I think it is important to locate the grammar of pain in the context of the lexicogrammar as a whole, to see it as an aspect of the overall construal of experience. Whether by analysing the grammar we could in any way contribute to the practical alleviation and management of pain I do not know. It might seem odd even to raise such a possibility. But I do believe that in order to understand any complex aspect of the human condition it is helpful to think about it grammatically. The boundary between the semiotic and the material worlds is by no means totally impermeable.
>
> (Halliday, 1998: 2)

Unlike Halliday, I am not interested as such in pain *per se* (for examples of work which examines this concept and how it is represented in language, see Semino, 2010, 2011;

Semino et al., 2017) However, I am interested in the linguistic encoding of 'felt' or affective experience (e.g. mental illness) and the encoding of the experience of mental illness in language. Because of this, I want to explore Halliday's final point in more detail, specifically the idea that the material world (what we can take to mean illness of some sort for our purposes here), and the semiotic world (what we can take to mean language). In answer to the question Halliday indirectly raises, I do not believe that prescribed changes in the language we use to describe illness may alleviate that illness to some extent (if this were the case then linguists may be more popular than they currently are). However, the notion that language and our experiences of the material world (specifically, in this case, having mental illness) can affect each other forms the theoretical basis for many mental illness anti-stigma campaigns that prescribe particular linguistic forms over others. For example, the *Time to Change* campaign launched in the UK in 2007 by the mental health charities Mind and Rethink Mental aims to end stigma around mental illness. On the *Time to Change* website is a resource called "Mind Your Language' which includes media guidelines on how to report on mental illness responsibly. The 'Mind Your Language' pages advise journalists to avoid referring to people with mental illnesses as 'patients' and instead use 'service-users'. Furthermore, they advise not to use the word 'suffer' in the context of mental illness, e.g. 'person suffering from mental illness', instead advising journalists to write 'person experiencing mental illness' (Time to Change, 2019). Although *Time to Change* do not state it explicitly on their website, these prescriptions about language use are based on the ideological content of 'suffer' and 'experience', which most speakers of English would agree convey different meanings. However, unsurprisingly, no linguistic evidence to support these claims is cited on the *Time to Change* website. In the following sections, I explore the grammatical form and ideological function of 'suffer' vs. 'experience' in order to explore whether there is linguistic evidence to support these linguistic prescriptions.

8.4. On the grammar of 'suffering'

In the previous section, I discussed the ideological weight of using certain words over others. Yet there is a quandary in that the report of the linguistic analysis that has to be conducted in order to explore that ideology inevitably requires words which may be ideologically loaded themselves. This is that the description of the semantic (and ideological) content of words necessarily requires words. This quandary is well documented by semanticists talking about the metalanguage of linguistic analysis (de Swart, 1998). In linguistic analysis of the sort conducted in this chapter, the metalanguage needed may present a further issue which is that the words used to describe ideology in a language may also carry ideological content. What I mean by this is that it is not the case that I can discuss the language representing the process of experiencing/suffering/having mental illnesses neutrally because my metacomment on the language may include ideologically loaded words. Indeed, Halliday (1998) writes of the process of 'having' pain that through using 'have' or 'has', "the grammar sets up a structural configuration of possession (process type "relational: possessive"). Some person […] becomes the owner of this thing" (Halliday, 1998: 4). For this reason, in contrast to the other chapters in this thesis where I have described people's 'having' of mental illnesses as 'experiencing' mental illnesses, I will refer to the process of having mental illness as just that - 'having'- in order to distinguish the concept of 'having' mental illness from my analysis of how that 'having' is representing through transitivity processes in the language, although I do note that even this most basic description of 'having' mental illness contains ideology. In this section I look at how the process of 'having' mental illness is represented through verb processes in the most prototypical texts for each year and in the MI 1984-2014 corpus more generally. I start by exploring the verb 'suffer'. To serve as a reminder of the process types in transitivity analysis, I reintroduce Figure 4.2 from Chapter 4 below (Table 8.1).

Process	Type	Subcategories	Elements
Material	'doing'	Intention (MAI) Supervention (MAS) Event *(if inanimate actor)* (MAE)	Actor, Goal (optional)
Verbalization Process	'saying'	(VP)	Sayer, Verbiage, Goal
Mental	'sensing'	Cognition (MC) Reaction (MR) Perception (MP)	Senser, Phenomenon
Relational	'being'	Intensive (RI) Possessive (RP) Circumstantial (RC)	Carrier, Attribute

TABLE 8.1. TYPES OF PROCESSES IN A TRANSITIVITY ANALYSIS (ADAPTED FROM HALLIDAY, 1973, JEFFRIES, 2010 AND SIMPSON, 1993)

Of the most prototypical texts, 11 out of 31 include the verb 'suffer' in the context of a person having a mental illness, with some articles using the verb up to three times. The frequency of 'suffer' as a verb in the prototypical text sample is indicative of the high frequency of 'suffer' in the corpus more generally (10,925 instances; 186.59 pmw). In addition to these verbal usages of 'suffer' in the prototypical text sample, it is also important to note that nominalised forms of the verb are also present; for example, 'care given to sufferers' (1987), 'anxiety sufferers' (1996), and 'bulimia sufferer'(2008). Whilst these usages are interesting and will be mentioned in passing in this chapter, the naming strategies that journalists use to refer to people with mental illness, such as 'sufferer', are discussed in more detail in Chapter 7. Table 8.2 shows instances of the verb 'to suffer' in the prototypical texts sample.

DATE	Element	Process	Element
1984	**73 per cent of women sent to Holloway prison's psychiatric unit, C1 wing, during that time** had already been diagnosed as	**suffering**	**from mental illness**
1985	He cites as an example **the current psychiatric fashion of diagnosing young black people as who present disturbing behaviour** as	**suffering**	from **a 'cannabis psychosis'**
1988	**Mr Dukakis**	**suffered**	**a deep depression**
1992	**'She** was	**suffering**	**the sudden worsening of depressive illness**
1999	**A DRUG DEALER who stabbed a man to death in an argument over £500 of cannabis** was […] Trelfa began	**suffering** **suffering**	from **clinical depression** **from depression**
2001	I've known **people** who've	**suffered**	**with depression**
2002	Scientists from Tilburg University, in Holland found **women**	**suffering**	**severe attacks**
2005	Psychiatrists agreed that Eltom was Eltom had already	**suffering** **suffered**	from severe depression, a bout of severe depression
2008	I don't think we have any idea how many men	**suffer**	with Bulimia
2012	he diagnosed Bonser as	**suffering**	from a personality disorder
2013	Last week the inquest into tragic Linzi Mannion showed she had been […] **Emma Cadywould**, who threw herself in front of a train after […] More than 15% of **new mothers**	**suffering** **suffering** **suffer**	from postnatal depression what the coroner called **one of the worst cases of postnatal depression he had seen** with **this illness.**

TABLE 8.2. EXAMPLES OF 'SUFFER' IN THE PROTOTYPICAL TEXTS SAMPLE

As would be expected from 'suffer' in this context, where the experiencer of the suffering is animate (e.g. new mothers, women, people), all the examples relate to mental processes of perception (MP) or processes that "are best described as states of mind or psychological events" (Bloor & Bloor, 1995: 116). The elements in mental processes are Senser and Phenomenon, as detailed in Table 8.3:

Participant: Senser	Process: Mental	Participant: Phenomenon
She	(had been) suffering	(with) this illness

Participant: Phenomenon	Participant: Senser	Process: Mental
The bulimia	she	(had been) suffering (from)

TABLE 8.3. EXAMPLE PARTICIPANTS AND PROCESSES IN MENTAL PROCESSES

Matthiessen & Halliday write of mental clauses that "a mental clause construes sensing […]" and includes "a process of consciousness involving a participant endowed with consciousness" (Matthiessen & Halliday, 1997: 18). I will return to the semantics of 'suffer' and the process types it can be coded as in Section 8.5 '*Experiencing* symptoms and *suffering* from illnesses'. First, however, I will discuss 'experience' and 'suffer' in the particular context of Table 8.2.

Table 8.2 shows that 'suffering' from mental illness is never a phrase used by a person with a mental illness in the sample; instead, 'suffering' is attributed to others, e.g. 'new mothers suffer with this illness'. This is interpretatively interesting because it removes any agency from the discursively created 'sufferer' because people typically suffer from things outside of their control. Furthermore, R1 noun collocates of 'suffer' (those collocates that appear one word to the right of the node word) in the BNC demonstrate that 'suffer' is a negatively valenced word. The collocates of 'suffer' include *damage, injury*, *pain*, *loss*, *harm* and *torture*. A concordance analysis of 'suffer from' in the BNC attests the interpretation that people typically suffer from things outside of their control.

innocent families having to	**suffer from**	that
people who make the laws aren't the people who	**suffer from**	them
we have to	**suffer from**	such incompetence
Bangladeshi communities	**suffer from**	the effects of poor housing
unable to replace management, the company may	**suffer from**	inept leadership

TABLE 8.4. CONCORDANCES FOR THE PHRASE 'SUFFER FROM' TAKEN FROM THE BNC

In the instances of 'suffer from' in the BNC shown in Table 8.4, the people and entities 'suffering from' something are all doing so as a result of someone or something out of their control. This interpretation is supported by Shweder who writes "to suffer is to experience a disvalued and unwanted state of mind, body or spirit" (Shweder, 2003: 76). What is more, it appears in the examples that the result of the action that led to suffering (e.g. being unable to replace management, the creation of laws), is that the people affected have little or no ability to rectify or deal with the thing that caused the suffering. For example, bad management has caused inept leadership and there is no possibility of rectifying the bad management situation, and the public have little or no say in changing the laws passed by people who do not represent them. What these instances of 'suffer from' in the BNC illuminate about the instances of 'suffering' in the prototypical texts sample, then, is that the choice to represent the process of having mental illness as 'suffering' creates the discursive role of the agentless 'sufferer' in the texts. Moreover, the instances of 'suffer from' in the BNC show that there is a precedent for 'suffer from' to convey a sense in which the 'sufferer' is precluded from dealing with the cause of their suffering.

The use of 'suffering' over another candidate term to convey the process of having mental illness may seem appropriate in some of the extreme cases in the prototypical texts. For example, an article from 2014 describes a new mother jumping in front of a train, because arguably, 'experiencing' does not convey the severity of her mental condition in the same way that 'suffering' does. However in some of the texts

in the sample, an alternative way of representing having mental illness may well have been possible without permuting the meaning in the same it would have in the 2014 article; as, for example, in an article from 2001, whose wider context reads '*I've known people who've suffered with depression and have had to take anti-depressants*'. In this example, it is clear that the people with depression are, at least to some degree, in control of and managing their depression, as they have sought medical intervention and are taking medication to control it. In this case then, 'experiencing' rather than 'suffering' may be argued to better represent the process for the reasons outlined above, where 'suffering' seems to suggest a lack of ability to deal with the cause of the suffering (recall that the media guidelines suggest using the verb 'experiencing' over 'suffering' when describing mental illness). In replacing 'suffer' with 'experience' in this case, the process stops being one where the agency of the actor is removed and instead becomes a process that is consciously experienced, and which creates the Senser the discursive role of 'experiencer'. Changing 'suffer' to 'experience' not only reinstates agency to the Senser but also recasts the illness (in these cases, depression) as a Phenomenon that is experienced (and then dealt with) and not something that is 'suffered' from indefinitely because the actor has no agency.

A concordance analysis of verbal usages of 'experience' in the BNC shows that the lexical item 'experience' conveys the sense of 'learned', i.e. through experiencing something, a person is better equipped to deal with it (or other related things) at a later date. In contrast, 'suffer' cannot encode this sense because it removes the agency of the Senser, and, as previously stated, there is precedent in the language more generally for 'suffer' to carry a sense of powerlessness to deal with a negative cause. Table 8.5 shows concordances of the verb 'experience' in the BNC. The table shows that unlike 'suffer', 'experience' is neither negatively nor positively valenced, but does convey a sense of actively living through something and is bounded, i.e. temporally fixed. For example, in (1) the experience lasts for the amount of time the visitors are in the park; in (2) the experience is caused by, and lasts as long as a particular sporting

event; and in (3) the experience is temporally bound by the amount of time spent visiting a tourist attraction.

1	Visit Murtonpark to	**experience**	the joys of the countryside
2	I was the first one to	**experience**	this sort of pressure, while players like Ian Wright and Keith Curle
3	Discover the answers as you	**experience**	for yourself the sights and sounds of their daily lives
4	Disabled visitors may	**experience**	difficulties due to gravel surfaces, slopes and steps.
5	long-lasting relief from sensitive, embarrassing external itching you can	**experience**	at any time of the month
6	as many as 61% of women	**experience**	it and suffer a certain amount of discomfort as a result.

TABLE 8.5. CONCORDANCE FOR THE VERB 'EXPERIENCE' IN THE BNC

Further to exploring the usage of 'experience' as a verb in the BNC, an analysis of 'experience' as a noun in the BNC attests the argument that experience implies 'learned' through examples such as '*He was in great pain but he used his skill and experience to escape*', '*with 24 years experience behind us we know what it takes to make an event stand out*', and '*that person has gained the skills and experience to get another job more easily*'. Further evidence for the interpretation that 'suffer' creates powerless participants in a way that 'experience' does not is that a statistically significant collocate of 'victim' is 'suffer' (MI = 7.14), whereas 'experience' (whilst still a statistically significant collocate) collocates much less frequently with 'victim' (MI = 5.62). This is an interesting finding because the media guidelines outlined by *Time to Change* suggest that journalists should avoid using 'victim'. This association between 'victim' and the verb 'to suffer', then, suggests that it could be the case that an increase

in frequency of the verb 'experience' may not only be a more appropriate way to describe the process of 'having' mental illness (because it suggests that the person is more in control of their illness than 'suffer' suggests and is more positively valenced), but may also result in a decrease in the use of 'suffer', which would also result in a decrease in the usage of other related problematic forms identified by *Time to Change*, such as 'victim'.

To return to the media guidelines published by *Time to Change*, I will now investigate whether 'suffering' is a verb used by people with mental illness in the corpus. I do this (1) to explore my earlier observation that all the instances of 'suffer' in the sample are used in reference to a person with mental illness and not by the person 'having' the mental illness, and (2) because if 'suffer' is negative in the way that *Time to Change* suggest it is, then we can expect people not to use 'suffer' to describe their own experiences with mental illness. To do this I searched the corpus for [first person pronoun + was suffering from/was suffering with/am suffering with/was suffering from]. Table 8.6 shows a summary of my findings with example concordances.

Phrase	Raw freq.	Freq. pmw	Example concordance
I was suffering from	240	4.1	it was by chance that I discovered **I was suffering from** anxiety attacks
I was suffering with	12	0.02	the doctor decided **I was suffering with** depression.
I am suffering from	42	0.72	My guilty secret is that **I am suffering from** mental illness.
I am suffering with	3	0.05	**I am suffering with** anorexia, binge-eating, purging, laxative abuse and exercise

TABLE 8.6. INSTANCES OF [FIRST-PERSON PRONOUN + SUFFER] IN THE MI 1984-2014 CORPUS

Table 8.6 shows that people with mental illness do refer to their 'having' of mental illness as 'suffering'. However, whilst this finding gives an indication that 'suffering' is used self-reflexively, it shows that 'I + [the verb to suffer]' is much less common in

the corpus than 'suffering' more generally (the lemma 'suffer' as a verb occurs 46,546 times in the corpus without "I" within 5 words right and left of it). The low frequency of first-person accounts from people with mental illness is due in part to the underrepresentation of first-person accounts in the corpus more generally. However, a search of [I + experience] indicates that there are other ways of referring to having mental illness in the corpus using the first-person pronoun. A concordance search shows that although first person voices in the corpus are few, examples of people talking about 'experiencing' mental illness are present. Table 8.7 shows examples of each usage.

Phrase	Raw Freq.	Freq. pmw	Example concordance
I was experiencing	70	1.2	After my daughter was born, **I was experiencing** post-natal depression and instead of cutting I burned myself, deliberately, with paint-stripper.
I am experiencing	8	0.14	It often means that **I am experiencing** difficulties with my life that I wasn't aware of and I need to deal with my feelings, rather than push it down with food and anaesthetise myself.
I experienced	166	2.84	It's been so long since **I experienced** these things that it took me a while to catch on that I was actually depressed, as opposed to suffering an iron deficiency or being slothful.
I experience	40	0.68	**I experience** all the high highs and the low lows.

TABLE 8.7. INSTANCES OF [FIRST-PERSON PRONOUN + EXPERIENCE] IN THE MI 1984-2014 CORPUS

Taken together, Tables 8.6 and 8.7 show that, in terms of relative and raw frequency, there is minimal difference between whether first-person accounts describe the

process of having mental illness as 'suffering' from a mental illness (total relative frequency of all forms 4.89), or 'experiencing' a mental illness (total relative frequency of all forms 4.86). However, 'experience' as a verb (10,925 instances; 186.95 pmw) is much less frequent in the corpus than 'suffer' (50,404; 860.87 pmw) overall. This means that that although [I + 'suffer'] is the more frequent of the two forms, [I + 'experience'] is proportionally four times more frequent in the corpus that [I + suffer][73].

In addition to frequency information, Tables 8.6 and 8.7 reveal a qualitative difference in how 'suffer' and 'experience' are used that indicates that unlike 'suffer', 'experience' foregrounds the person's individual experiences of mental illness such as discovering that their symptoms constituted a mental illness and their reaction to their feelings during this process. The concordances of 'suffer' reveal no such description of experiences; rather, 'suffer' appears to only relate to a diagnosable illness as a complete thing (as opposed to a collection of symptoms that one experiences). I will explore this conceptual difference between 'suffer' and 'experience' in Section 8.4.

Another interesting finding from Tables 8.6 and 8.7 is that there is a preference in first-person usages of 'suffer' and 'experience' for the process to be described in the past tense (i.e. *was* suffering with/from, experienc*ed*/*was* experiencing), with 488 of the total 581 instances referring to the past tense. The preference to refer to mental illness in the past tense in the corpus is perhaps unsurprising, given that the majority of newspaper reports discuss events that have already happened; however, arguably, the preference for representing mental illness as occurring in the past does not accurately represent the reality of many mental illnesses that occur throughout peoples' lives. The tendency for people to refer to having mental illness in the past provides further support for the argument that 'experience' is a more fitting way to

[73] I arrived at this number by dividing the raw frequency of [I + verb] (here 'experience' or 'suffer') by the total number of instances of 'experience' or 'suffer' as a verb in the corpus. I then multiplied the number by 100 to get a percentage (suffer = 0.6%, experience = 2.6%) and then divided the percentage of 'suffer' by 'experience'.

describe having mental illness, as 'experience', unlike 'suffer' was shown to refer to bounded phenomena in Table 8.5.

Tables 8.6 and 8.7 reveal that 'suffer' occurs slightly more frequently in the corpus with the first-person singular pronoun in raw terms of raw and relative frequency; however, 'experience' is proportionally more frequent. This suggests that [I + 'experience'] is the more common form overall. The fact that there is a tendency for first-person accounts of having mental illness to not use 'suffer' is further attested by the fact that in one first-person account in Table 8.6 which does included [I + 'suffer'], this is contained within indirect reported speech:

> the doctor **decided I was suffering with** depression.

In the example, [I + 'suffer'] is a process attributed by another actor, e.g. a medical expert, in which the 'I was suffering with depression' clause constitutes the grammatical direct object of a sentence in which the grammatical subject (the doctor) 'decides' the person is suffering. The MI 1984-2014 corpus provides further evidence for the finding that many instances of [I + suffer] are reported speech. Table 8.8 show a sample of instances in which [I + suffer] is presented as indirect speech or free indirect speech (Leech & Short, 2007; McIntyre et al., 2004):

doctor told me	**I was suffering from**	anxiety neuroses and offered me Valium
I started having palpitations and the **doctors said**	**I was suffering from**	panic attacks.
Then **he explained** simply and rationally that	**I was suffering from**	a bipolar illness.
I explained my symptoms to a **student nurse** who I had just met and she **said** it sounded like	**I was suffering from**	panic attacks

He **told me**	**I was suffering from**	depression and that I was shocked and frightened
My GP said	**I was suffering from**	an acute stress reaction
He said	**I was suffering from**	acute anxiety and ECT
I know that **my doctor says** that	**I am suffering with**	severe shock

TABLE 8.8. CONCORDANCE FOR [I AM SUFFERING] IN THE MI 1984-2014 CORPUS

Although we cannot know how faithful the reported speech is (i.e. whether the health professional actually said what they are reported to have said), Table 8.8 shows that the apparent frequency of reflexive [I + suffer} is less than Table 8.6 suggests. This finding gives evidence for the notion that people with mental illness do not refer to their experience of it as 'suffering'. Interestingly, a concordance search of [said I was experiencing] (as opposed to [said I was suffering]) in the corpus yielded just one hit from 2011 which reads:

> I went to my doctor and said "something is not right" and **he said I was experiencing** postnatal depression again".

What the low frequency of [said I was experiencing] indicates is that, at least in the corpus, 'suffer' is a more common way to refer to having people having mental illness by mental health professionals.

Interestingly, this finding is in keeping with a theme in the corpus more generally which is that 'experts' are the people in the corpus that 'say', 'think' and 'find' things, whereas people with mental health illnesses are constructed as the people things are said about, thought about and found out about.

8.5. 'Experiencing' symptoms and 'suffering' from illnesses

So far in this chapter I have presented evidence to suggest that the mental processes 'experience' and 'suffer' encode different meanings. I have argued that the two processes are different in the way that they encode temporality, i.e. 'experiencing' is bounded whereas 'suffering' is not. I have also argued that concordance evidence from the BNC shows that 'suffer' occurs in negative contexts, whereas 'experience' does not appear to occur in a wholly positive or negative context (see Tables 8.4 and 8.5). Another difference that was revealed in the analysis conducted in Section 8.4 is that 'suffer' appears to remove agency from the Senser in a way that 'experience' did not.

In this section I explore the lexicogrammatical differences between 'experience' and 'suffer'. The reason for doing this is that in addition to the differences between 'experience' and 'suffer' listed above, an interesting observation to come from the concordance analysis of 'suffer' and 'experience' in Section 8.4 was that there appears to be a conceptual difference between 'experiencing' mental illness and 'suffering' from mental illness in the corpus. I argued that this conceptual difference appears to be related to the nature of the Phenomenon in the process. Specifically, I raised the possibility that 'suffer' appears to relate to diagnosable illnesses in the majority of cases, whereas 'experience' appears to relate to symptoms in the majority of cases. I will explore this in more detail now.

A concordance analysis of sentences in the corpus that contain both 'experience' and 'suffer' as verbs suggests that this conceptual difference provides further evidence for the thesis that whether 'suffer' or 'experience' is used is related to diagnostic status; i.e. if a person has symptoms but no diagnosis of mental illness then the process is described as being one of 'experiencing'. In contrast, where there is a diagnosis (or where there is a reference to a diagnosable mental illness, as opposed to a set of symptoms (e.g. 'depression' rather than 'low mood, fatigue and insomnia') then the process is described as one of 'suffering'. Of course, there are instances in the corpus

where 'suffer' and 'experience' are used as synonyms, presumably to provide linguistic variety in the newspaper article; however, an analysis of the 639 instances (10.91 pmw)[74] in the corpus indicates that there is a pattern in which having symptoms of an illness is described as 'experiencing', and living with a diagnosable condition is described as 'suffering'. This is curious when presumably the symptoms a person experiences constitute the illness and therefore experiencing those symptoms is the cause of the suffering; e.g. suffering from depression means also suffering from low mood. Table 8.9 shows a series of concordances attesting this interpretation. Note that depression can be both a symptom and a diagnosable condition in itself.

men **suffering** from **postnatal depression** **experience** **paranoia**, **delusions** and, in some cases, **thoughts of suicide**
mothers **suffering** from **postnatal illness** will be **experiencing** a sustained and **pervasive depression** and **lowness of mood**
People who **suffer** from **phobias** are afraid of the **feelings** they **experience** when they get **anxious**
If you are **experiencing** these **symptoms** in combination you could be **suffering** from **depression** and should seek medical advice
Many of those who **experience** **panic attacks** also **suffer** from **depression**
A further 213 claimed they had **experienced** **suicidal thoughts** and 407 said they were **suffering** **depression**
Mr Morrison was found to be **suffering** from **chronic schizophrenia** after **experiencing** a **delusion** that he was being shot at with poison darts
There are a range of support mechanisms for people **suffering** from **depression** and we strongly recommend anyone **experiencing** **signs of depression** consults an appropriate health professional
She was **suffering** from an **acute psychotic episode** despite never previously **experiencing** **mental illness**
People who **suffer** from **SAD** **experience** a **lack of energy, have sleep problems and mood changes** and they feel **anxious** and have **difficulty concentrating**

TABLE 8.9. CONCORDANCE ANALYSIS OF SENTENCES CONTAINING THE LEMMAS 'EXPERIENCE' AND 'SUFFER' IN THE MI 1984-2014

[74] These 639 instances were returned from the following search: [lemma search 'suffer' within 10 words L&R of lemma 'experience']

What this tendency suggests is that once something is named as a diagnosable illness, the process of 'experiencing' becomes one of 'suffering' (even if, prior to diagnosis, the person was experiencing all the symptoms of an illness anyway). This finding is an interesting one because it raises the question of why it should be the case that 'suffer' relates to diagnosis but 'experiencing' the symptoms does not. I argue that a plausible reason for this conceptual difference is that along with a diagnosis comes a diagnostic label (e.g. 'a schizophrenic', 'a bulimic'), and, as previous research has attested, those labels are stigmatising because such labels define the person by their illness. Indeed, the American Psychiatric Association (APA) recognised that the use of labels such as 'schizophrenic' in which the person is described by their illness "had the potential to promote bias, devalue others, and express negative attitudes" (Granello & Gibbs, 2016: 31; APA, 1992; Haghighat & Littlewood, 1995).

An analysis of nouns that form the subject collocates of the lemmas 'suffer' and 'experience' reveals that labels that define a person by their illness are more likely to occur with 'suffer' than 'experience'. There are no instances in the MI 1984-2014 corpus in which a subject collocate of 'experience' is a label that defines a person by their illness.

Subject collocate of 'suffer'	MI score
victim	7.14
schizophrenic	4.97
alcoholic	4.24
anorexic	4.19

TABLE 8.10. SUBJECT COLLOCATES OF THE LEMMA 'SUFFER' IN THE MI 1984-2014 CORPUS.

In addition to the APA's identification of the problems associated with using language that defines people by their illness, *Time to Change* also state that such forms should be avoided. Knowing this then, Table 8.10 reveals again that 'suffer' is not just a problematic form on its own, but is also associated with problematic forms like 'schizophrenic', 'anorexic' and 'sufferer' (recall also that 'suffer' was identified as a collocate of 'victim' in Section 8.4). 'Suffer' as a verb, then, is not just problematic

because it misinterprets many peoples' experiences of mental illness (because it is possible to live well with a mental illness); arguably 'suffer' also provides a linguistic trigger for stigma creation and perpetuation in language.

In section 8.4, I described 'suffer' in the prototypical text samples as a mental process (MP) and said I would return to the semantics of 'suffer' in this section. I will now summarise my findings so far and discuss them in relation to the transitivity categories outlined in the introduction to this chapter.

So far in this chapter I have argued that 'suffer', unlike 'experience', occurs in contexts in which the Senser (i.e. the person suffering) has little or no agency in controlling their 'suffering'. I evidenced this claim by showing concordances of 'suffer from' in the BNC (see Table 8.4). Furthermore, in this section I have argued that there is a conceptual difference between 'suffer' and 'experience' which results in 'suffer' being used when referring to a diagnosable illness, and 'experience' being used when referring to symptoms. Consider, for example, 'John suffered from bipolar disorder' and 'John experienced mood swings, low mood and delusions'. In this section I will examine in more detail the semantic and syntactic properties that I argue underlie the conceptual difference between 'experience' and 'suffer'. I will do this by looking in specific detail at invented sentences to explore the semantic and syntactic possibilities of 'suffer' and 'experience'.

As previously stated in Section 8.4, 'suffer' in the examples shown in Table 8.2 is a mental process of perception, a process which requires a conscious and animate Senser and a Phenomenon (a state of mind or psychological event). Just like 'suffer', 'experience' is also a mental process. An indication that both 'suffer' and 'experience' are mental processes is that 'experience' can be used (in SFG terms) in the same way that 'suffer' can be used. Note that in terms of syntax, the two do not function in the same way because 'experienced with' is syntactically ungrammatical. Examples (A) and (B) show this:

(A) He **suffered** with depression [Senser + Mental Process + Phenomenon]

(B) He **experienced** depression [Senser + Mental Process + Phenomenon]

However, there is a difference between 'suffer' and 'experience' that is not evident from examples (A) and (B). This difference is that unlike 'experience', which is always a mental process, 'suffer' can also be a material process (recall that material processes describe "doings or happenings" [Matthiessen & Halliday, 1997: 17] and have an Actor + Material Process + Goal configuration). As described in Table 8.1, material processes can be further subdivided into material supervention processes (hereafter MAS) where the process is unintentional (e.g. 'the woman fell'), material action events (hereafter MAE) where the actor is inanimate (e.g. 'time ran away with me'), and material intention processes (hereafter MAI) where the actor is animate and the process is intentional (e.g. 'the woman wrote her thesis'). The reason that 'experience' cannot be a material process is because the process of 'experiencing' necessarily requires a conscious entity, but 'suffering' does not. Furthermore, unlike 'suffer' used in a mental process which is intransitive (i.e. does not require an object), the 'suffer' used in a material process is transitive; that is, it requires an object. Compare, for example, (A) and (B) with (C) and (D):

(C) The fence **suffered** damage from the wind [Goal + MAE + Circumstance + Actor]

(D) Ben **suffered** a broken leg [Goal + MAS + Circumstance + ∅Actor]

In (C) the process 'suffered' is an MAE process because the Actor (the fence) is inanimate; in (D) the process 'suffered' is MAS because the Actor (Ben) is animate and the process is one that does not encode intention, i.e. people do not suffer broken legs intentionally. What (D) shows is that 'suffer' can function in a very similar way to the example shown in (A), but as a material action. If we try to do the same test for 'experience', we see that it cannot be a material action. See for example, (E):

(E) The fence **experienced** damage from the wind*[75]

I argue that in (E), 'experience' is semantically incongruous (and therefore semantically "unacceptable" if we are to borrow the parlance from syntax research – see footnote 75) because fences cannot 'experience' – they do not have consciousness as they are not animate.[76] The point of showing these examples here is that I argue that they show that 'suffer' has very close contextual associations (as is the case in (D)) with mental processes which are in turn associated with inanimate, non-conscious actors with no agency; for example, 'the fence' in (C). In terms of the ideological effects of using 'suffer' over 'experience', then, it could be argued that 'suffer' can be used as a device to represent the 'having' of mental illness in a way that reduces the person's agency in mental processes because of its association with non-conscious things without agency[77].

In addition to there being a conceptual difference between 'suffer' and 'experience', however, there also appears to be a further conceptual difference between 'suffer' and 'suffer from'. For example, consider the semantic differences between [suffer + preposition] (i.e. where suffer is intransitive) and [suffer + noun] (i.e. where suffer is transitive) in examples (F) – (I):

[75] As is the convention in syntax research, I use an asterisk to denote that the sentence described is unacceptable. However, in this context the acceptability is based on semantic in/congruity and not syntactic in/congruity.

[76] The reader may at this point be thinking about other examples where 'experience' is acceptable with inanimate actors; for example, "Flight 123A experienced extreme turbulence". In cases such as this one, I would argue that the nominal group constituting the Actor (the Flight 123A) is a metonymic substitution and actually refers not to the plane but the passengers. As such, I think my point above is still sound.

[77] The reader may argue that 'suffer' can also appear in material actions of intention processes, such as "she didn't suffer fools", and therefore my point about 'suffers'' association with agentless actors may be weakened because in this context, 'suffer' does encode intention. However, I argue that this is an idiomatic usage and is therefore an exception.

	Sentence	Syntactic Properties	Proposition
(F)	**Ben suffered *a broken leg***	Transitive Verb + NP (MAS)	Ben sustained a broken leg
(G)	Ben **suffered *with a broken leg***	Intransitive Verb + PP (MP)	Ben suffered with the effects of his broken leg, e.g. his (mental) experience of dealing with a broken leg
(H)	The table **suffered *a broken leg***	Transitive Verb + NP (MAE)	The chair sustained a broken leg
(I)	The chair **suffered *with a broken leg****	Intransitive Verb + PP (no transitivity classification)	Semantically incongruous – a chair cannot suffer with something

TABLE 8.11. INVENTED SENTENCES FOR [SUFFER + PP] AND [SUFFER + NP]

Table 8.11 shows that in the cases where the verb 'suffer' is followed by a preposition, the thing(s) being suffered from relate to mental experiences or affective states (e.g. requiring some cognition) because the meaning of (F) is distinct from (G). The reason for this, then, must be the addition of a preposition. Example (I) provides further evidence for this interpretation as (I) is semantically incongruous, which suggests that 'suffering with' cannot be a process attributable to an inanimate object. Put simply, this finding suggests that 'suffer + preposition' indicates a mental process. In light of this finding, consider examples (J) and (K) below, which relate to suffering in a mental health context:

(J) Ben **suffered *with*** depression

(K) The bulimia she **suffered *from***

Like the majority of the instances of 'suffer' in the prototypical text sample, examples (J) and (K) show a preference for describing the suffering of mental health conditions

as 'suffering from' something rather than 'suffering' something, e.g. 'suffering from bulimia' and not 'suffering bulimia'. What the findings in Table 8.11 suggest is that it is not just that there is a tendency in the language for diagnosable illnesses to be 'suffered' rather than 'experienced'; in addition, there also appears to be a tendency for linguistically encoded mental illnesses to be 'suffered from' and physical illnesses to be 'suffered'. This is an interpretively interesting tendency because it suggests that there is not just a conceptual distinction between 'suffer' and 'experience' but also a conceptual difference between mental 'suffering' and the 'suffering' caused by physical things. This could be taken as linguistic evidence for a perceived duality between physical/mental illnesses, i.e. physical illnesses are 'suffered' and mental illnesses (or the mental effects of physical illness) are 'suffered from'. Examples (F) and (G) support this interpretation as do the R1 noun collocates of 'suffer' and 'suffer from' in the BNC (see Table 8.12).

R1 noun collocates of 'suffer' in BNC			R1 noun collocates of 'suffer from' in BNC	
1	fools		1	asthma
2	injury		2	chronic
3	damage		3	stress
4	pain		4	malnutrition
5	loss		5	severe
6	death		6	lack
7	torture		7	mental
9	harm		8	having
8	withdrawal		9	animosity
9	losses		10	incontinence

TABLE 8.12. R1 NOUN COLLOCATES OF 'SUFFER' AND 'SUFFER FROM' IN THE BNC.

Table 8.12 shows that the hypothesis that mental suffering is 'suffered from' and physical suffering is 'suffered' appears to bear out in the language generally. For example, the things being 'suffered' in the BNC have a tendency to relate to physical things such as 'damage', 'loss' and 'torture'. In comparison, things that are described as being 'suffered from' appear to have a tendency to relate to non-physical things,

e.g. 'stress', 'mental' and 'animosity'. It is the case, of course, that mental and physical 'causes' can both result in mental illness, and therefore what I present here is simplistic, but Table 8.12 and the examples given above do, I argue, provide evidence for the *perception* that there is a duality. Moreover, it is important to state that this lexicogrammatical quirk is just that – a quirk; it is a tendency rather than a rule and there are examples in the corpus that contradict this interpretation; for example, "the frequent result is that the wife suffers depression' (1987), "he still suffers depression, but rarely acutely" (2002). However, a concordance search for 'suffers depression' (40 instances; 0.7 pmw) vs. 'suffers from depression' (310 instances; 5.29 pmw) reveals a marked tendency in the corpus to refer to 'suffering' using the latter structure.

We may now consider where 'experience' fits here. To do this, I will summarise my findings of 'suffer', 'suffer from' and 'experience' so far. My findings include that 'experience' is not necessarily positive or negative, whereas 'suffer' does occur in negative contexts. Furthermore, 'experience' necessarily requires a conscious entity in order to be semantically congruous, whereas 'suffer' does not (note that grammatically a non-conscious entity can 'experience' things). I have also argued that, unlike 'suffer' which is not temporally fixed (rather it appears to relate to a state of indefinite 'suffering'), 'experience' does appear to be temporally fixed. Finally, I argued that there appears to be a tendency in the language generally for 'suffer' to occur with physical things, whereas 'suffer from' occurs with non-physical things.

Taking these findings together, we may categorise the features of 'experience', 'suffer' and 'suffer from' in a feature matrix as follows:

experience	suffer	suffer [+ prep]
+ animate	+/- animate	+/- animate
– negative	+ negative	+ negative
+ mental	+/- mental	+ mental
+ bounded	– bounded	+ bounded

TABLE 8.13. FEATURE MATRIX FOR 'EXPERIENCE', 'SUFFER' AND 'SUFFER + PREPOSITION'

One of the reasons that Table 8.13 is illuminating (other than the fact that it allows for the description of how 'suffer', 'experience' and 'suffer from' are used in the corpus and in general language corpora) is that it provides a more nuanced meaning of the verb 'suffer'. For example, the Oxford English Dictionary (OED) lists several senses of 'suffer' which are related by identical etymologies (from Latin *sub-ferre*). The grammatical difference between the senses of 'suffer' are to do with whether the verb takes a grammatical object. For example:

Sense	Definition	Example usage
Suffer, v. (transitive)	To undergo, endure	"Every one who does wrong is to suffer punishment by way of admonition."
Suffer, v. (intransitive)	To undergo or submit to pain, punishment, or death.	"She was suffering from what she was pleased to call a fit of depression."

TABLE 8.14. SENSES, DEFINITION AND EXAMPLE USAGES FOR 'SUFFER' (TAKEN FROM THE OED)

Table 8.14 shows that there is little discernible difference between the senses of 'suffer' (intransitive) and 'suffer' (transitive). For example, both describe suffering as being concerned with enduring things and undergoing things. Knowing this, I argue that the analysis in this chapter has consistently shown (in the MI 1984-2014 corpus and the BNC) that there is a difference between '[suffer'+ noun] (transitive) and [suffer + prep] (intransitive), namely that [suffer + prep] has a tendency to occur in contexts in which the suffering relates to a non-physical thing (i.e. mental), e.g. 'suffer from depression', whereas ['suffer + noun] occurs in material acts, e.g. 'suffered a broken

leg'. As such, the analysis I have conducted here provides a more nuanced understanding of the verb 'suffer'. Moreover, what I have shown here builds on previous research in linguistic anthropology and linguistics on the semantic content of 'suffer', such as that of Shweder who states that "suffering is a state of mind" (Shweder, 2003: 76), and that of Wierzbicka (2016) who in her explication of the differences between 'suffering' and 'pain' states that "suffering implies consciousness whereas pain does not" (Wierzbicka, 2016:29). My findings suggest that 'suffering' is not always state of mind as Shweder (2003) suggests (because inanimate objects can suffer things), and does not necessarily imply consciousness (as Wierzbicka (2016 suggests) for the same reason. It is the case, however, that ['suffer' + prep] is a mental state and implies consciousness because 'suffer from' usually occurs in a mental process, i.e. with a Senser and Phenomenon.

By way of an overview of the syntactic features of '[suffer'+ noun] (transitive) and [suffer + prep] (intransitive) described above, and the semantic features described in the feature matrix (Table 8.13), Table 8.15 provides a delineation of a set of lexicogrammatical heuristics:

SENSE 1: Suffer, v. transitive			
[subject = animate; object = concrete] → event is material and bounded Examples: *Ben suffered a broken leg* *Ben suffered a twisted ankle* *Ben suffered a fall** NB: in some cases where an NP complement of a PP is bounded the resulting clause can be semantically marked, e.g. 'Ben suffered from a ball to the face', however if the temporally bound event may have lasting consequences then it is acceptable, e.g. 'Ben suffered from a blow to the head' * semantically marked		**subject = animate; object = abstract] → event is mental and bounded** Examples: *Helen suffered intense pain* *Helen suffered toothache* *Helen suffered embarrassment*	**subject = inanimate; object = concrete] → event is material and bounded** Examples: *The car suffered a serious engine failure* *The fence suffered wind damage* *The house suffered water damage* N.B. As stated earlier, inanimate objects cannot appear with abstract subjects e.g., the fence suffered embarrassment*
SENSE 2: Suffer, v. intransitive			
[subject = animate; object = concrete] → event is mental and unbounded Examples: *Ben suffered from a broken leg* *Ben suffered from a twisted ankle* *Ben suffered from a fall*	**[subject = animate; object = abstract] → event is mental and unbounded** Examples: *Helen suffered from intense pain* *Helen suffered from toothache* *Helen suffered from embarrassment*	**[subject = inanimate; object = concrete]→ event is material and unbounded** Examples: *The car suffered from a serious engine failure* *The fence suffered from wind damage* *The house suffered from water damage*	**[subject = inanimate; object = NONE] → event is mental and unbounded** Examples: *John suffered* *It suffered* *The car suffered** * semantically marked

TABLE 8.15. LEXICOGRAMMATICAL HEURISTICS FOR 'SUFFER'

Taking all the information detailed in Table 8.15 together, it makes sense that a diagnosable condition, which is by its very nature unbounded (i.e. not temporally fixed; rather, boundless), will appear with 'Sense 2' (intransitive 'suffer'). In contrast, symptoms, which may be bounded (e.g. 'suffered from a bout of depression'), or may not be bounded ('suffered a bout of depression') may appear with both senses.

In this section, I have discussed the conceptual differences between 'experience' and 'suffer'. I argued that this difference was to do with the diagnostic status of the Senser in the clause. Specially, I argued that the corpus revealed a tendency to refer to 'experiencing' symptoms of an illness whereas diagnosable conditions (such as, for example, schizophrenia) are described as being 'suffered' from. I also showed that, unlike 'experience', 'suffer' collocates with labels that both the American Psychiatric Association and the anti-stigma initiative *Time to Change* have both identified as problematic ways to refer to people with mental health, such as 'schizophrenic' and 'anorexic'. Furthermore, I have argued that there is also a lexicogrammatical difference between 'suffer' in intransitive and transitive contexts. Specifically, I argued that there is a tendency in the corpus, and in general language corpora, for ['suffer' + noun] to refer to suffering from physical things (e.g. a broken leg), and for ['suffer' + prep] to exhibit a tendency to be used to refer to affective states of suffering mentally. I argued that this tendency could provide some linguistic evidence for the notion that there is a perceived duality between physical/mental illnesses.

In the next section, I bring together the findings of this chapter and offer some further points for discussion.

8.6. Conclusion

I started this chapter by introducing transitivity analysis and its utility for analysing ideology in texts. I also introduced Halliday's (1998) argument that a language is a theory of human experience and that "in order to understand any complex aspect of

the human condition it is helpful to think about it grammatically" because "the boundary between the semiotic and the material worlds is by no means totally impermeable." (Halliday, 1998: 2). I used Halliday's point about the link between the material and semiotic worlds (*viz.* the social world and language respectively) to argue that it was this link that formed the basis of many anti-stigma initiatives that prescribe language use in order to change the way mental illness is discussed in the media. I then explored the verb 'suffer' which I identified as a frequently used verb in the sample of prototypical texts (occurring in 11 out of 31 texts). I analysed the lexicogrammatical function of 'suffer' and showed that 'suffer' was a mental process of perception in the sample taking a Senser, Process, Phenomenon configuration. I showed that in the sample from the MI 1984-2014 corpus, and in general language corpora, 'suffer' occurred in contexts in which the Senser has little or no control over their suffering, and that suffering was negative. I then explored the verb 'experience' because this word has been identified by the UK anti-stigma initiative *Time to Change* as a less stigmatising way to describe having mental illness than 'suffer'. My analysis of 'experience' showed that 'experience', unlike 'suffer', did occur in contexts where people had control over the thing they were experiencing. Furthermore, I argued that the discursive role created by 'experience', i.e. 'experiencer', encoded greater agency than the discursive role created by 'suffer', i.e. a sufferer. I also showed that 'suffer' collocates with other problematic labels such as 'schizophrenic' and 'anorexic' (which identify a person by their illness), as well as 'victim'. An analysis of self-reflexive uses of [I + suffer] vs. [I + experience] also reveals that 'suffer' is proportionally less likely to be used in first-person narratives because 'suffering' is attributed to people with mental illness by others e.g. medical professionals, in reported speech. I also argued that my corpus and the BNC reveal a conceptual difference between 'suffer' and 'experience' that is concerned with the diagnostic status of the Senser in the process. Specifically, I found that 'experience' was more likely to occur in relation to symptoms, whereas 'suffer' was more likely to occur where the Sensor has a diagnosis of an illness. Further to this, I argued that the MI 1984-2014 corpus revealed that there

is not just a conceptual difference between 'experience' and 'suffer' but also 'suffer' and 'suffer with/from'. I explored this difference and demonstrated that there is a tendency in the MI 1984-2014 corpus and the BNC for 'suffer' (intransitive) to occur in contexts where the suffering is material (i.e. physical), while 'suffer' (transitive) has a tendency to occur in mental processes and relate to affective states, e.g. mental suffering.

I did this analysis to answer the research question '*What processes are associated with mental illness?'* What my analysis has found, then, is that 'suffering' is a frequent and salient process associated with mental illness, whereas 'experience' is less common. Despite this, a linguistic analysis of the instances of 'experience' in the corpus reveal it to better represent the reality of having a mental illness because it encodes agency and consciousness and implies learnedness. Moreover, I have shown that the way 'suffer' is used (i.e. intransitive = material suffering vs. transitive = mental suffering) reveals linguistic evidence for a perceived duality between mental and physical suffering, and by extension, mental and physical illness.

In Chapter 5, I reviewed the existing literature on the representation of mental illness in the press. I argued that whilst there are many studies investigating stigma in mental illness reporting, those studies do not account for how stigma is discursively created. Often, stigma in the previous literature was predicated on the topic matter of the articles as a whole, such as whether the articles described mental illness in the context of criminality or violence. Clearly, this macro, thematic analysis of stigma in mental illness reporting is useful and provides insight into the salient features of the stigmatisation process in articles reporting on mental illness. However, none of the previous research provides a comprehensive account of how language has the potential to shape our perception of the world by creating and perpetuating stigma. Moreover, to the best of my knowledge, the *Time to Change* media guidelines (e.g. avoid 'suffer' and instead use 'experience') are not based on any linguistic research, and as a result the suggestion that 'suffering' is more problematic than 'experiencing' must be based on dictionary definitions and intuitive responses to these words, rather

than the analytical findings of a linguistic study. Recent research that forms part of the *Time to Change* campaign from 2016 does feature references to 'pejorative language' in their coding schema as a 'stigmatising element' (Rhydderch et al., 2016); however, this term is never explained, and no part of the research describes what constitutes pejorative language, or how the research team coded for it. Existing research commissioned by Time to Change, then, is thematic (in that the analysis takes place at the level of the article rather than at the level of the word/clause/sentence) and is primarily quantitative[78].

In bringing attention to this fact, I do not wish to suggest that I disagree with Time to Change's assessments about 'suffer' or replacing 'suffer' with 'experience'. I believe it is a sensible intuition and one that I agree with as a user of the language. What I do believe, however, is that linguistic research should underpin such prescriptions. I argue that the research I have conducted here into the choices that journalists make when representing the process of experiencing mental illnesses as 'suffering' provides this because it offers evidence that supports Time to Change's suggestion that 'experiencing' is a more appropriate form than 'suffering'. I argue that the findings reported in this chapter, for example the collocational associations between words like 'suffer' and 'schizophrenic', 'anorexic' and 'victim', are evidence of the pernicious and subtle textual associations that create stigma. Micro linguistic analysis, then, provides a method that is more nuanced than thematic analysis of the kind reported in the existing research. Moreover, nuanced textual analysis of this kind is precisely what Halliday describes when he writes that the experiential basis of language is concerned with "patterns of meaning that are installed in the brain" (Halliday, 2003: 15). Knowing this, the contribution of this chapter is not only to give greater insight into salient processes associated with mental illness, but also to illuminate nuanced textual associations that could be said to create stigma.

[78] This is something that the research manager at *Time to Change* is aware of, and *Time to Change* are interested in different methods (*Time to Change*, personal communication, September 26, 2017).

In the next chapter I explore whether news articles accurately portray the symptoms of mental illnesses.

9. Do newspaper reports accurately represent the symptoms of mental illness?

9.1. Introduction

So far in this thesis, I have explored the terms 'mental health' and 'mental illness' and showed how their meanings have changed over the time period covered by the corpus. I have also demonstrated the utility of linguistic analysis for revealing ideological meaning through naming practices in Chapter 7 and through transitivity analysis in Chapter 8. These previous chapters have exemplified a bottom-up approach in that they have let the data reveal how mental illness is discussed in the press. In these chapters, I have made very little reference to the medical context or reality of mental illness, as my aim was merely to describe the way in which mental illness is discussed in the corpus. For example, I have reported that particular naming strategies appear to be more closely associated with particular illnesses, but I have not spent a great deal of time dealing with whether this association is grounded in the reality of the illness, e.g. whether the illness affects these particular people more than others. In this chapter, then, I will depart from a purely linguistic analysis and move on to an analysis that is informed by the medical context in which these illnesses lie. In doing so, I address the following research question listed in the introduction to this thesis:

5. Is the depiction of mental illness realistic?
    5.1. Are the symptoms of each disorder type (e.g. depressive illnesses) accurately portrayed in the press?

My intention is that the findings in this chapter will be useful to linguists interested in mental health discourse and practitioners engaged in the field of mental health alike. Moreover, I hope that these findings will begin, in some small sense, to provide insight into which areas to target to make the reporting of mental illness more realistic.

However, before I conduct my analysis, it is worth explaining my use of the terms 'realistic' and 'accurate' in relation to the reporting of the symptoms of mental illnesses. The first point to note is my own understanding of these terms in this context. I take 'realistic' and 'accurate' to refer to the how closely the press report the symptoms as stipulated by *Mind* and by the *NHS*; that is, if a press report of a mental illness includes the list of symptoms of that illness given by *Mind* or the *NHS* then it is a realistic report of that illness. Interpreting and analysing 'realistic' in this way provides parameters for my analysis; however, I am aware that the clinically recognised symptoms (i.e. those symptoms listed by the *NHS*) are general and therefore imperfect. For instance, it is important to note that there is sometimes a distinction between being scientifically accurate and reflecting a person's lived experience. That is, a person with a mental illness may not experience that illness (i.e. through the symptoms they experience) in the way described by clinicians. Nonetheless, in order to ensure the replicability and falsifiability of my analysis, I have chosen to use the symptoms listed by *Mind* and the *NHS*. Furthermore, this analysis constitutes one of the first explorations of the representation of symptoms in news reports. As such, my analysis is intended to provide a baseline measure of how the press discuss the symptoms of mental illnesses. As Filer (2019: 7) writes

> "it would be a grave mistake to dismiss any of this as unimportant. Yes, it's a dispute about language, but in the mad, mad, world of mental healthcare *language is everything*. A simple truth […] is that the overwhelming majority of psychiatric diagnosis aren't arrived at by looking at blood tests or brain scans or anything of the sort. Rather, it is the *words* people say – or do not say – as interpreted by professionals, that as much as anything else will determine a diagnosis
>
> (Filer, 2019: 7)

In addition to the points above about being realistic, we need also to consider the extent to which news reporting can be expected to be 'accurate'. By 'accurate' I mean the degree to which a newspaper represents the symptoms (recall that by this I mean *clinical* symptoms) of mental illnesses. Although there are constraints on news reports, such as article length and news values (Bednarek & Caple, 2017), it is not unreasonable to expect journalists to observe a baseline level of accuracy in their reports due to the fact that part of the purpose of the press is to inform. It is also important to note that being accurate is not simply about what is included in an article, but also what is routinely omitted.

The way I bring together the medical knowledge of mental illnesses and language about mental illnesses is via two analytical processes, each containing several steps. The processes are designed to include linguistic analysis that is enriched by information about the real-world context of mental illness; in particular, the statistics about, and symptomology of, particular mental illnesses. Combining these processes allows for the exploration of the linguistic representations of mental illness (i.e. whether the symptoms of a particular illness are represented accurately). The reason for exploring symptoms specifically is that the inaccurate representation of symptoms of mental illnesses in the press has been identified by Wahl et al. (2002) as

a means of contributing to stigma around mental illnesses. In their research, Wahl et al. (2002) analysed 300 articles (collected using the generic search term 'mental illness') to identify various themes that contributed to whether mental illness reportage had changed over time (specifically whether it has become less stigmatised). Wahl et al. (2002) noted that despite the significant body of research into negative portrayals of mental illness (e.g. associations between mental illness and violence or criminality) some aspects of mental illness reportage had been neglected in the previous literature. They write

> little attention has been paid in previous research as to whether or not psychiatric labels are explained or defined (for example providing information about the symptoms or behaviors that characterize the mental illness mentioned
>
> (Wahl et al. 2002: 13)

Despite identifying this gap in the existing research, Wahl et al. (2002), only analysed whether any symptoms were present in the articles (a simple yes or no on their coding scheme) and not what constituted the symptoms, or whether the symptoms were accurate or not. As a result of this methodological decision, Wahl et al. (2002) were only able to report that it was rare for news articles to include symptoms of mental illnesses. The analysis conducted in this chapter, then, can enrich Wahl et al's (2002) study in two ways. First, the data is more specific than the data Wahl et al. (2002) used (i.e. it is illness-specific) which means that the data I analyse in this chapter is more likely to contain references to symptoms because it contains articles on specific illnesses and not articles reporting on mental illness generally. Moreover, the MI 1984-corpus is larger and more representative of mental illness discourse than Wahl et al's (2002) data. Second, the corpus linguistic methods I use in this chapter are more systematic and rigorous than the thematic analysis conducted by Wahl et al. (2002) who rely on the interpretation qualitative data. Adopting methods from (corpus)

linguistics (e.g. keyness analysis) allows for the comparison of particular features of the language (e.g. whether a series of texts discusses a particular symptom) across large datasets

Table 9.1 outlines the analytical processes I use to explore whether the symptoms of mental illnesses are (accurately) represented in the MI 1984-2014 corpus.

Process 1: **Exploring the social reality of illness(es)**	**Process 2:** **Exploring the linguistic manifestation of illness(es)**
1. Generate list of mental illnesses in culture under investigation (here Anglo-western, British) 2. Group illnesses into disorder types (e.g. personality disorders or trauma disorders informed by DSM-V) 3. Check symptomology for each illness type (informed by *Mind* A-Z of Mental Health	1. Compile subcorpora to target specific mental illnesses (done using method outlined in Chapter 5) 2. Use corpus linguistic analytical tools to: a. Generate a keyword list and key semantic domain list for each illness subcorpus to see whether the specific corpus contains words in keeping with symptoms b. Compare findings from steps 2a with the most prototypical text in the corpus c. Use concordance analysis using syntactic search frame [query term and…] to see what symptoms, or other mental illnesses the query term is being textually equated with d. Conduct sketch thesaurus search to compare findings

TABLE 9.1. PROCESS OUTLINE FOR COMPARING THE SOCIAL REALITY OF THE MENTAL ILLNESS DISCOURSE DOMAIN WITH THE LINGUISTIC MANIFESTATIONS OF THE DISCOURSE DOMAIN

The rest of this chapter will be dedicated to working through the processes outlined in Table 9.1.

9.2. Grouping mental illnesses

In Chapter 5, I introduced *'Figure 5.1 Lexical items grouped by disorder type as categorised in DSM-V'* that detailed how the mental illnesses covered in the corpus are grouped according to disorder type. I compiled the figure using a range of information I collected from the *Mind A-Z of Mental Health* and the description of the illnesses listed in the DSM-V. A slightly amended version of this figure is shown below (Figure 9.1). It is important to note that the groupings of the illnesses do vary according to which source you consult. For example, *Mind* describe OCD as an anxiety disorder, whereas DSM-V describes it as an obsessive compulsive disorder. I have taken the grouping listed by DSM-V in the majority of cases. The variation in the groupings serves as evidence that greater transparency is needed in how mental illnesses are named (and therefore the associations they trigger in the minds of the people reading literature on them). Moreover, such groupings have changed over time and continue to change between the editions of the DSM. One example of this is PTSD (posttraumatic stress disorder) which was, until DSM-V, listed as an anxiety disorder not a trauma disorder (McNally, 2009). With this variation in mind, the categories listed in the DSM-V and the *Mind A-Z of Mental Health* serve as rough groupings rather than exact categories. Moreover, researchers of mental health and illness have to be mindful of the unfixed and culturally specific nature of mental illness. *Mind* write on the subject of describing mental health problems in different cultures:

> Different cultures have different approaches to mental health and mental illness. Most western countries agree on a similar set of clinical diagnoses and treatments for mental health problems. However, cultures in which there are other traditions or beliefs may not use these terms[.]
>
> (Mind, 2019d)

Knowing this, all groupings and linguistic descriptions of mental disorders (i.e. labels) are imperfect (cf. the research reported in Chapter 2 by van Os (2016) who argues that

the label 'schizophrenia' doesn't exist[79]). This notwithstanding, the categories outlined in Figure 9.1 do provide a reflection of the different mental disorder types used by medical professionals, and as a result, offer a starting point from which to analyse the linguistic patterns within each disorder type. It should be noted that any grouping of this kind is always a simplification of the whole picture (where many illnesses share symptoms with others and many people who experience a specific mental illness will also experience other mental illnesses). In addition to showing the different groups of mental disorders and how specific illnesses fit within them. Figure 9.1 details the illness subcorpora I use in this chapter to target texts discussing disorder types. For example, depressive disorders include depression, seasonal affective disorder (SAD) and postnatal depression (PND); however, I only used the depression subcorpus as the target corpus. The reason for this decision is a practical one: Wmatrix (Rayson 2008), the corpus tool I use in this chapter to conduct keyness analyses has a limit on the size of corpus that can be uploaded to the tool. Wmatrix's size limit precluded me from merging some of the corpora (e.g. the Depression corpus and the SAD corpus) because together the two would exceed Wmatrix's limit. Moreover, the illness subcorpus creation procedure was designed to group illnesses and not to split them. For this reason, some illness corpora include all the distinct illnesses within the disorder type. An example of this is the BipolarDisorder Corpus which contains articles reporting on bipolar disorder, hypomania and hypermania. The reason for this is that hypermania and hypomania are both symptoms of bipolar disorder in addition to being separate illnesses. Furthermore, the illness subcorpus creation procedure outlined in Chapter 5 means that it is reasonable to expect that articles on specific illness within a disorder type (e.g. SAD and PND) would be included in the more general corpus (e.g. depression). Appendices 9.3 and 9.4 provide information about

[79] van Os's point is that schizophrenia, unlike other psychotic disorders, is given its own label. The labelling of schizophrenia and not other psychotic disorders suggests that schizophrenia is a discrete illness when he argues it is not.

the illness subcorpora, what search terms each illness subcorpus contains and the workflow procedure for the illness subcorpora compilation

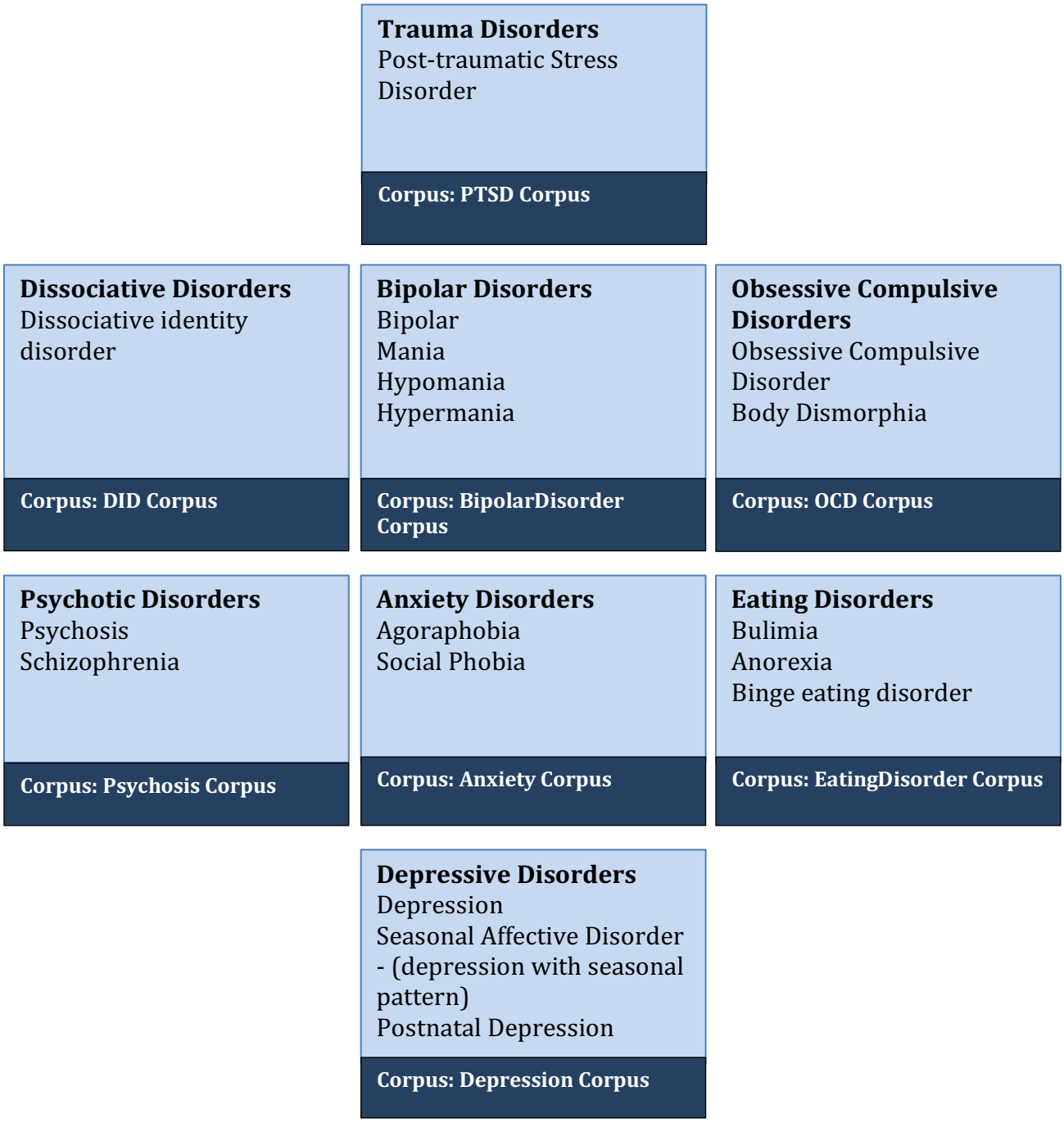


FIGURE 9.1. ILLNESSES GROUPED BY DISORDER TYPE AS CATEGORISED IN DSM-V WITH RELEVANT TARGET CORPUS.

Figure 9.1 shows the list of illness contained in the corpus grouped by disorder type. For example, bulimia, anorexia and binge eating disorder are all categorised as eating disorders. Figure 9.1 provides us with the information needed to fulfil steps A and B in Process 1 (outlined in Table 9.1). The next step is to gather information about the symptoms of each illness. For clarity, Process 1, Step 3 will be completed by disorder type (i.e. in disorder group).

Sections 9.3 – 9.10 will explore the how the symptoms of each illness or disorder type are represented in the corpora. These sections will describe the symptoms of each illness contained within each disorder type using keyness analysis as an indication of the symptoms covered in the corpus, before moving on to the linguistic representation of each illness type, i.e. through linguistic analysis Steps 2a and 2b. Section 9.11 will bring together analyses conducted in sections 9.3-9.10 and will explore whether the keyness analysis findings are supported by collocation analysis. Section 9.10 will be concerned with Steps 2c and 2d in Table 9.1. I do not analyse autism spectrum disorders in this chapter because I am not concerned with autism as a condition in itself; the Autism corpus was created purely as a means of collecting data pertaining to mental illnesses due to the higher incidence of mental illness in people with autism. Furthermore, I do not analyse personality disorders in this chapter due to there being noise in the corpus which prohibits keyness comparisons.

9.3. Trauma disorders

In this section, I discuss the top keywords and key semantic domains in the PTSD corpus in relation to the symptoms of PTSD. Trauma disorders are categorised by symptoms including flashbacks, intrusive thoughts, irritability, aggressive behavior and being jumpy or easily startled. PTSD is often caused by some form of trauma such as a car crash, being attacked or being in combat and can be described as 'delayed-onset PTSD' (where symptoms occur after six months of the trauma), 'complex

PTSD[80]' (where the trauma happened at an early age or lasted for a long time) and 'birth trauma' (where the PTSD is caused by a traumatic childbirth). Often people with a diagnosis of PTSD also experience other mental illnesses including anxiety disorders, depression and dissociative disorders (*Mind*, 2019e) As a result of the links between PTSD and these other mental disorders, it may be the case that the language used to describe PTSD in the PTSD corpus overlaps somewhat with the language used to describe these other illnesses. Furthermore, PTSD UK (2019) state that "[A]nyone can be diagnosed with PTSD, and it's estimated that 1 in 10 people develop PTSD. 1 in 5 firefighters, 1 in 3 teenagers who have survived a horrific car crash, 70% of rape victims, 2 in 3 prisoners of war, 40% of people who experienced a sudden death of a loved one, and an estimated 10,000 women a year following a traumatic childbirth, develop PTSD". Knowing this, we can expect that the possible causes of PTSD that are represented in the corpus will be varied.

The first steps according to Process 2 are to conduct keyness analysis at the level of the lexical item (keywords) and at the semantic level (key semantic domains), as such analyses make it possible to ascertain whether the symptoms of PTSD are represented in the language used about PTSD. The reason for such an analysis is to ascertain whether the depiction of symptoms is realistic in the corpus. Exploring whether the symptoms of illnesses are present and realistically described is important because arguably, a realistic depiction or description of an illness requires some discussion of symptoms to a) properly convey the experience of PTSD, and b) convey information to the public about the condition in order to help people who may be living with the condition (a realistic depiction of the symptoms of the illness allows people to determine whether they or someone they know may have it).

[80] *Mind* (2019) state that *complex PTSD* is a new term and some professionals instead refer to this type of PTSD as 'enduring personality change after catastrophic experience' (EPCACE) or 'disorders of extreme stress not otherwise specified' (DESNOS).

The keyness analysis I conduct in this chapter uses corpora (rather than individual texts) as the target and reference unit for comparison. Due to this, the keyness analysis reported is an abstraction from the individual articles (i.e. keyness results are based on statistical significance from the data as a whole and not on the individual texts). In order to address any potential disparity between the results of the keyness analyses and the content of the individual texts, I enrich my keyness analysis with qualitative analysis using the most prototypical text in the PTSD corpus to ascertain whether the keyness findings are supported by qualitative evidence. In order to do this, I use ProtAnt (Anthony & Baker, 2015) which is a tool that finds the most prototypical text (here the individual newspaper article) in a collection of texts (here the relevant illness subcorpus) based on the number of keywords each text contains compared with a reference corpus. Using a combination of keyness analyses and whole texts in my analysis (as opposed to relying wholly on keyness analysis) also counters the possible over-reliance on researcher-created units of analysis (i.e. corpora) at the expense of naturally occurring data (texts), a practice which researchers have previously argued can cause analysts to 'miss the trees for the forest' (Egbert & Schnur, 2018: 159). Furthermore, there is precedent for using prototypicality as a means of capturing texts that are representative of particular language feature or topic in discourse analytic research (Anthony & Baker, 2015) and research into news discourse (Bednarek & Caple, 2017: 146).

Table 9.2 shows the top 30 keywords in the PTSD corpus compared with all the other illness subcorpora minus the "MentalIllness' corpus. The reason for not including the MentalIllness corpus in the keyword comparison is that the MentalIllness corpus is a general corpus of mental illness in which articles that reference mental illness or mental health generally are. Due to the fact that I wanted to compare specific illness subcorpora with specific illness subcorpora (and therefore increase the possibility of that I retrieved more specific keywords), I opted to not include the MentalIllness corpus. Table 9.3 shows the top 30 key semantic domains in the PTSD corpus compared with the 'MentalIllness' corpus. All the key semantic

domain analyses I conduct in this chapter compares the target corpus with the MentalIllness corpus due to the fact that merging all the other illness subcorpora (i.e. all illness subcorpora minus the relevant target corpus) would exceed the upload limit on Wmatrix. For this reason, I made the decision to base all key semantic domain comparisons on the MentalIllness corpus because it was the most general (i.e. most likely to include a variety of lexical items related to all illnesses) and it did not exceed the token limit on Wmatrix.

Rank	LL.	Item	Rank	LL	Item
1	31832.062	ptsd	16	3779.657	MOD
2	14392.202	traumatic	17	3586.087	personnel
3	13329.759	veterans	18	3442.73	soldier
4	13329.759	soldiers	19	3430.118	falklands
5	8854.257	stress	20	3370.299	flashbacks
6	7644.844	war	21	2638.748	disorder
7	6790.572	iraq	22	2571.671	troops
8	6343.375	army	23	2564.377	armed
9	6235.9	military	24	2418.257	nightmares
10	5761.355	afghanistan	25	2418.257	served
11	5597.484	post	26	2025.045	ex
12	5011.571	combat	27	1986.226	serving
13	4611.952	trauma	28	1749.6	gulf
14	4045.618	servicemen	29	1620.687	service
15	3812.534	forces	30	1534.86	ministry

TABLE 9.2. TOP 30 KEYWORDS IN THE PTSD CORPUS COMPARED WITH ALL OTHER ILLNESS CORPORA MINUS THE 'MENTALILLNESS' CORPUS CALCULATED USING ANTCONC (ANTHONY, 2017)

Rank	LL	Domain	Rank	LL	Domain
1	21852.70	WARFARE, DEFENCE AND THE ARMY; WEAPONS	16	254.35	PEOPLE: MALE
2	1505.56	SAD	17	249.42	EVALUATION: BAD
3	1128.89	PRONOUNS	18	199.78	TEMPERATURE: HOT/ON FIRE

4	968.26	ANATOMY AND PHYSIOLOGY	19	195.23	DRINKS AND ALCOHOL
5	851.17	GEOGRAPHICAL NAMES	20	174.94	SAILING, SWIMMING, ETC.
6	801.97	MOVING, COMING AND GOING	21	167.46	VIOLENT/ANGRY
7	715.65	ANTI-WAR	22	163.90	UNMATCHED
8	628.38	DAMAGING AND DESTROYING	23	154.66	TIME: ENDING
9	461.55	KNOWLEDGEABLE	24	129.43	LINEAR ORDER
10	436.06	FLYING AND AIRCRAFT	25	128.73	TIME: LATE
11	364.11	PSYCHOLOGICAL ACTIONS, STATES AND PROCESSES	26	111.96	HINDERING
12	361.36	WORRY	27	100.80	PERSONAL RELATIONSHIP: GENERAL
13	344.80	FEAR/SHOCK	28	93.51	EMOTIONAL ACTIONS, STATES AND PROCESSES GENERAL
14	337.56	VEHICLES AND TRANSPORT ON LAND	29	88.95	TRYING HARD
15	311.03	SENSORY: SIGHT	30	84.59	TIME: PERIOD

TABLE 9.3. TOP 30 POSITIVE KEY SEMANTIC DOMAINS IN THE PTSD CORPUS COMPARED WITH THE 'MENTALILLNESS' CORPUS (LL CUT OFF 10.83, MIN. FREQ. 5, P < 0.001, *P* < 0.001) CALCULATED USING WMATRIX (RAYSON, 2008)

Tables 9.2 and 9.3 show that the symptoms of PTSD are well-represented in the PTSD corpus. Keywords such as 'flashbacks' and 'nightmares' indicate that the most salient symptoms of PTSD are being represented. Furthermore, symptoms such as anxiety and aggression are also represented in the key semantic domain analysis where 'worry' and 'violent and angry' are both statistically significant. Furthermore, the symptom of self-destructive behavior in people with PTSD is represented in the corpus as is evidenced by the DRINKS AND ALCOHOL semantic domain which contains

instances in which the newspapers report on the misuse of substances such as alcohol to deal with PTSD.

The keyness analyses of the PTSD corpus indicates that the most salient symptoms of PTSD are being represented in news reports, which is a positive thing. However, the symptoms of PTSD (and by extension the depiction of PTSD more generally) in the news as shown in the keyword and key semantic domain analysis indicates that PTSD is often only discussed in reference to a very specific group of people – veterans or ex-soldiers.

The major theme of war in reference to PTSD is also shown in the most prototypical text taken from the PTSD corpus which is a *Guardian* article on the topic of PSTD (specifically, criminal cases of Vietnam veterans in the USA where PTSD has been given as a mitigating circumstance). The most prototypical text in the PTSD corpus features a list of symptoms like those generated in the keyness analyses, as the extract below shows. It also indirectly references symptoms such as self-destructive behaviour and aggression.

> Servicing in Vietnam became a major defence in criminal cases. Post-traumatic stress disorder it is called, PTSD. At least 250 veterans accused of serious crimes have received a more lenient sentence by claiming to be PTSD sufferers - with nightmares, depressions, sleep loss and flashbacks touched off by sights, sounds or smells that remind them of Vietnam.
>
> (*The Guardian*, November 17, 1986)

Whilst the extract above does include a list of PTSD symptoms, it does so in the context of criminal cases which is a fairly specific set of circumstances in which to talk about PTSD given that the whole article is about a Vietnam War memorial service held in Washington DC. As a result, it could be argued that the article foregrounds PTSD in criminal cases rather than as a condition that many people experience outside of such circumstances. Moreover, the journalist's choice to describe the defendants as

"claiming" to be experiencing PTSD implicates that there is a possibility that the defendants are using PTSD as an excuse. This choice of word is an unhelpful depiction of PTSD and its effect on people's lives as it questions the validity of the lived experience of PTSD. Moreover, it should be noted that the list of symptoms includes 'depressions' but not 'depression', which also suggests that, at least at this point in time (1986), depression as a diagnosable illness is not recognised.

Reference to the Iraq War and military action in Afghanistan in the corpus (as indicated by the keyword list) is unsurprising given the dates the corpus covers and the statistics PTSD UK report on PTSD prevalence in war situations (2 out of 3 prisoners of war); however, what is surprising is that very little newspaper coverage is given to the other common causes of PTSD such as (sexual) assault and childbirth. In fact, only one keyword in the top 30 could possibly relate to childbirth ('trauma') and on closer inspection only seven (including one duplicated article in a local newspaper) of 1084 instances of 'trauma' appear in the phrase 'birth trauma', and all of these seven instances are used in the charity name 'The Birth Trauma Association'. This means that just 0.6% of the total number of articles in the PTSD corpus discuss 'birth trauma' using that label. Furthermore, when looking in more detail at the keyword list, 'childbirth' and 'childbirths' are only keyword numbers 457 and 1566 respectively, and there are only 76 hits in the whole corpus for '*births'. The earliest of the seven hits for 'birth trauma' is in 2004 which indicates that the label was not applied to PTSD caused by childbirth before this time, or that it was not a recognised illness. Evidence from PTSD UK seems to support this observation. They state: "It has long been recognised that following a difficult childbirth some women may go on to develop psychological problems. However, it is only relatively recently that it has become accepted that women can develop Post Traumatic Stress Disorder (PTSD) as a result" (PTSD UK, 2019).

Another way that PTSD UK refer to birth trauma is 'PTSD after birth'. In order to ascertain whether birth trauma was being discussed in the corpus under this label, I searched 'after birth' which only yielded two hits that both refer to an experiment on

mice that aimed to explore a genetic basis for PTSD. This finding indicates that 'birth trauma' as well as 'PTSD after birth' are not reported in the corpus using those labels.

What this analysis of birth trauma in the corpus shows is that whilst the symptoms of PTSD are present in the corpus (flashbacks, nightmares, aggression, etc.) the symptoms are discussed in relation to specific circumstances, particularly combat situations. This means that the reality of PTSD is not accurately portrayed as the reportage overall neglects other circumstances from which PTSD arises, particularly childbirth. Moreover, PTSD in women is underrepresented overall as is demonstrated by the low frequencies for birth trauma and also by the fact that the semantic domain 'people: male' is statistically significant. This representation of PTSD as an illness that affects men more than women (as is suggested by the prevalence of male pronouns and reference to combat situations where the subject is male) is in contrast to statistics into the prevalence of PTSD in the UK where women are more likely to screen positive for PTSD than men (around 5.1% compared with 3.7% of men) (Baker, 2018: 8). The findings from the linguistic analysis, taken together with the statistical information about the reality of PTSD, are important for informing an accurate depiction of PTSD in the press. As it stands, there is a disparity between the reality of the condition and the representation of it in the press. Arguably, the main purpose of the press is to inform the public. Articles reporting on PTSD, then, need to discuss the symptoms of PTSD in situations other than combat or military service, otherwise the press are not accurately representing the reality of PTSD.

9.4. Dissociative disorders

In this section, I discuss the top keywords and key semantic domains in the dissociative identity disorder (DID) corpus in relation to the symptoms of dissociative disorders. Dissociative disorders are characterised by periods of dissociation wherein a person may be unsure who they are or may have different identities. There are many

types of dissociative disorder; however, for our purposes here, I only detail the symptoms of Dissociative Identity Disorder (DID), as this was the only dissociative disorder searched for during the corpus compilation process.

DID, sometimes called multiple personality disorder (MPD), is a condition that causes people who experience it to have severe changes in their identity. For example, people with DID may have different identity states that each have "different patterns of thinking and relating to the world" (*Mind*, 2019f). These different identities may be different ages, or genders, and some people who have DID may refer to themselves as 'we'. People who experience DID may have periods of amnesia, for example, when other identities are in control (*Mind*, 2019f). These identities are sometimes referred to as 'parts'. People who experience dissociation may also experience other mental illnesses such as borderline personality disorder (BPD), depression, anxiety and OCD. Like trauma disorders, DID may be caused by a traumatic event – and medical experts believe that particular traumatic events are more likely to cause DID than others. These include abuse and neglect either from a young age or that takes place over a long period, or if caregivers are dissociative themselves.

The DID corpus is the smallest of the illness subcorpora as it contains just seven articles. This means that DID (or more precisely articles that contain DID more than any other condition) make up just 0.02% (11,953) of all the tokens contained in the MI 1984-2014 corpus. As a result of the low frequency of tokens, it is necessary to be wary of the generalisability of the keyness analyses reported in this section. Nevertheless, working through the same processes with the DID corpus as I have done with the other illnesses provides a means of ascertaining how well-represented the symptoms of DID are.

As was the case in Section 9.3, I will now present the results of the keyness analyses and compare my findings with the most prototypical text in the DID corpus. Tables 9.4 and 9.5 show the keywords in the DID corpus compared with all the other illness subcorpora minus the MentalIllness subcorpora, and the key semantic domains

in the DID corpus compared with the MentalIllness corpus respectively. Due to the low frequencies in the DID corpus, there are only 26 key semantic domains.

Rank	LL.	Item	Rank	LL	Item
1	578.143	hopkins	16	133.905	aliens
2	479.981	ufo	17	129.605	dummy
3	442.69	kim	18	127.918	noble
4	439.007	dissociative	19	105.124	budd
5	410.8	cueller	20	102.386	alien
6	348.479	perez	21	102.213	artists
7	308.514	abduction	22	99.794	mulumba
8	247.292	linda	23	98.907	annoyingly
9	231.477	words	24	98.064	aimee
10	188.314	abductees	25	95.516	wreaks
11	178.575	paragraphs	26	88.003	contains
12	154.041	de	27	87.402	dan
13	149.691	sapir	28	86.844	procrastination
14	148.685	ufos	29	86.844	swathe
15	135.05	personalities	30	85.402	malingering

TABLE 9.4. TOP 30 KEYWORDS IN THE DID CORPUS COMPARED WITH ALL OTHER ILLNESS CORPORA MINUS THE 'MENTALILLNESS' CORPUS CALCULATED USING ANTCONC (ANTHONY, 2017)

Rank	LL.	Domain	Rank	LL	Domain
1	237.35	ARTS AND CRAFTS	16	19.81	MEASUREMENT: AREA
2	154.72	LANGUAGE, SPEECH AND GRAMMAR	17	19.66	MEASUREMENT: SIZE
3	112.66	PERSONALITY TRAITS	18	19.64	GEOGRAPHICAL TERMS
4	110.97	UNMATCHED	19	19.23	KNOWLEDGEABLE
5	93.56	EVALUATION: UNAUTHENTIC	20	16.91	PEOPLE: MALE
6	59.73	SEEM	21	16.48	GENERALLY KINDS, GROUPS, EXAMPLES
7	54.50	FLYING AND AIRCRAFT	22	16.34	QNATOMY AND PHYSIOLOGY
8	53.12	PERSONAL NAMES	23	14.51	RECIPROCAL
9	52.37	BEING	24	13.51	MENTAL ACTIONS AND PROCESSES
10	43.83	RECORDED SOUND	25	11.64	SPEED: FAST

11	43.73	COMPARING: DIFFERENT	26	11.34	ELECTRICITY AND ELECTRICAL EQUIPMENT
12	38.55	COMPARING: USUAL			
13	35.89	TIME PERIOD: SHORT			
14	27.43	COMPARING: UNUSUAL			
15	22.16	PAPER DOCUMENTS AND WRITING			

TABLE 9.5. TOP 30 POSITIVE KEY SEMANTIC DOMAINS IN THE DID CORPUS COMPARED WITH THE 'MENTALILLNESS' CORPUS (LL CUT OFF 10.83, MIN. FREQ. 5, *P* < 0.001) CALCULATED USING WMATRIX (RAYSON, 2008)

The keyword list in 9.4 is not particularly illuminating in terms of the symptoms of DID. This is likely due to skew in the corpus due to its small size which means that the keywords are heavily influenced by one long article (4963 words) contained in the corpus titled "They're coming: can space aliens really have abducted the former Secretary-General of the United Nations? It sounds absurd but thousands of Americans seem convinced by a rumour which has become a *cause célèbre* even outside the mad world of modern 'Ufology'", which was published in 1993 in *The Independent*. The article, as the title suggests, reports on the story that UFO enthusiasts (including two people called Budd Hopkins and Jay Sapir) believed that the then Secretary-General of the UN, Javier Perez de Cueller, had been abducted by aliens. In the article, there is a discussion that the symptoms of alien abduction may be mistaken for the symptoms of dissociative disorders (reported in the article as MPD, specifically). Clearly then, this article is not about DID, but rather alien abduction. Keywords that have been discounted due to noise in the corpus have been shaded light grey on Table 9.4 for clarity.

I conducted a concordance analysis of the remaining keywords in order to investigate them in more detail. They revealed that three articles in the corpus focused on particular people (hence the significant results for proper names in the semantic domain analysis and the proper names in the keyword list). Two of these stories related to an artist called Kim Noble who has a diagnosis of DID, but one of these

stories reports on the death of a teenager called Alex Mulumba who was stabbed to death in London in 2006 and was later found to be linked with a gang. As a result, the press coverage of Mulumba's death quickly turned from reports that foregrounded the tragedy of Mulumba's death and his father's heartbreak to those that foregrounded Mulumba's part in knife-related gang crime. In the newspaper article, which was published in *The Independent*, the journalist writes:

> it might also be because his son was as convincing in his role as a good son as he was in his role as a schoolboy gangster. I'm not quite suggesting that Alex Mulumba might have been suffering from what used to be called multiple personality disorder, and is now called dissociative personality disorder. But I'm convinced that the untouchable affectlessness that is reported among many of the young men on society's margins - and some of the young women -might be an indication that they may be suffering from a range of related dissociative disorders. The concept of dissociation is easy to grasp as being on a continuum reflecting a range of experiences from those which are quite normal: daydreaming, "switching off", to those which are extreme: blacking out, or feeling utterly alienated from one's own physical actions. It's worth noting that a number of the other activities that worry parents, such as playing video games or watching music videos for hours, are well-known to trigger "zoned-out" states. There's surely some mileage in the idea that "street culture" is dissociative, encouraging as it does a retreat into a simple, violent, two-dimensional world
>
> (*The Independent*, June 28, 2006)

In writing this, even though the journalist writes that they are "not quite" suggesting that Mulumba had DID, the fact that she mentions it at all has the effect (via a flout of the Gricean maxim of quantity) of linking Mulumba's ability to hide the different parts of his identity (i.e. the good son vs. the violent gang member) and dissociative disorders like dissociative identity disorder. This is a particularly unhelpful representation of DID as it places it in the context of violence (here gang violence) when there is no evidence that Mulumba had any dissociative disorder. Moreover,

speculating about somebody having mental illness is against *Time to Change's* media guidelines which stipulate "don't speculate about someone's mental health being a factor in the story unless you know this to be 100% true." (*Time to Change*, 2019a). Furthermore, the journalist neglects to offer a list of symptoms of dissociative disorders in context, or any follow up to the comments that show there is no basis for linking DID and violence. Instead all that is offered is a vague list of symptoms, e.g. being 'zoned-out' which is not descriptive enough to provide clear information to the public about DID.

The rest of the articles that focus on people are concerned with Kim Noble, an artist with DID. Both of the articles about Noble were published in local newspapers. Unlike the Mulumba article, both of the articles discussing Noble feature first-person narration from Noble in which she describes her own experiences of living with DID. The most prototypical text of the DID corpus is an article about Noble. Below is an extract which discusses Noble's role in a mental health initiative that showcases the work of artists that are staying or working in hospital:

> Kim has been diagnosed with dissociative identity disorder. She has no memory of the personalities who take over her body. But painting is a way for her to connect with them. "When we started painting, it was something I found we all had in common," said Kim, 52. "The personalities are all different - one's a Catholic, for instance, and one likes a drink. But with art we were on the same page. "If I lose time when I'm in the house on my own, I can go into my art room, see that somebody has painted, and know from the style who has been in. It helps me know more about them
>
> (*Kentish Weeklies*, October 17, 2013)

The article featuring the extract above does not include a thorough list of symptoms, but instead focusses on living with DID, particularly how arts initiatives may help those with mental illnesses. This is in line with the findings of research conducted by Atanasova et al. (2019) who found that recovery was a key theme in their research into press reports on mental health and arts initiatives. Although a list of symptoms is

absent in this article, overall the article can be said to provide a positive portrayal of DID because it features first-person narration of a person who is living well with DID and previous research into positive portrayals have identified first-person accounts from people with mental illness as contributing to a positive portrayal (e.g. Nairn & Coverdale, 2005; Wahl, 2003).

An earlier article published in the *Daily Mail* in 2011 also discusses how Noble is using art as a way to live with her DID. Moreover, the article talks about previous trauma in Noble's life that may have caused her dissociation. In addition to featuring first-person narration from Noble, the article includes a description of DID and its symptoms from a medical consultant. In his description of DID, the consultant lists the symptoms of DID and also provides information about how it is linked to PTSD. Arguably this is a positive description of DID and its symptoms as it provides an account of the lived experience of one person (Noble) as well as the general symptoms and causes of DID. As a result of this, the reader has access to an account of DID that shows a person living well with DID and is made aware of the symptoms and causes of DID that may fall outside of Noble's experience.

These articles indicate that overall in the DID corpus, when DID is discussed, symptoms or first-person accounts from people with DID are included. However, the low number of articles reporting on DID (just three in the whole corpus) suggests that there ought to be more coverage of DID in the press more generally given that Positive Outcomes for Dissociative Survivors (PODS) reports figures that state that between 0.4 and 3.1% of people who are not currently involved with mental health services have DID (PODS, 2019), which means that more people have DID than schizophrenia.

9.5. Bipolar disorder

In this section, I discuss the top keywords and key semantic domains in the BipolarDisorder corpus in relation to the symptoms of bipolar disorder. Bipolar

disorder is a mood disorder that causes people to have very high moods (mania or hypomania) and very low moods (depressive episodes). Often people with bipolar disorder experience extreme mood swings that affect their everyday life. Bipolar disorder was previously referred to as manic depression. The symptoms of bipolar, e.g. periods of mania or depression (sometimes called bipolar episodes), vary in length depending on the person and the specific diagnosis a person has, as medical professionals recognise three types of bipolar. These are bipolar I, bipolar II and cyclothymia and are diagnosed according to the severity of the moods (i.e. extreme mania vs. more extreme depression) the person experiences. The symptoms of mania include feeling euphoric, feeling like you cannot be harmed and feeling uncontrollably excited (*Mind*, 2019g). Mania can also include some psychotic symptoms such as hearing voices. Mania may result in people losing social inhibitions, spending money excessively and not sleeping. The symptoms of mania can last for over a week. Where people with bipolar disorder experience the symptoms of mania for a shorter period of time, or the symptoms are more manageable, the term used to describe the mania is *hypomania*. In contrast to mania, depressive episodes are characterised by feeling upset, tired and having low self-esteem. As a result of this, people experiencing a depressive episode may eat more or less than usual, may misuse drugs and may attempt self-harm or suicide (*Mind*, 2019g). Some people with bipolar disorder also experience mixed episodes where their mood may fluctuate between feeling depressed and manic or hypomanic over short periods of time.

Tables 9.5 and 9.6 show the top 30 keywords and key semantic domains for the BipolarDisorder Corpus.

Rank	LL.	Item	Rank	LL	Item
1	13680.657	bipolar	16	789.388	i
2	4182.797	disorder	17	787.405	condition
3	2278.026	manic	18	750.298	mania
4	1815.786	mental	19	685.082	ii
5	1538.836	zeta	20	636.855	swings

6	1423.199	catherine	21	564.451	carrier
7	1281.017	illness	22	533.304	danes
8	946.429	douglas	23	515.718	was
9	938.966	her	24	515.464	kerry
10	928.64	depression	25	502.977	episodes
11	889.329	jones	26	494.428	actress
12	852.96	diagnosed	27	448.606	he
13	838.638	mood	28	424.075	diagnosis
14	831.81	fry	29	424.071	highs
15	793.094	she	30	421.293	homeland

TABLE 9.5. TOP 30 KEYWORDS IN THE BIPOLARDISORDER CORPUS COMPARED WITH ALL OTHER ILLNESS CORPORA MINUS THE 'MENTALILLNESS' CORPUS CALCULATED USING ANTCONC (ANTHONY, 2017)

Rank	LL	Domain	Rank	LL	Domain
1	2801.47	pronouns	16	161.90	music and related activities
2	1906.60	comparing: different	17	160.64	linear order
3	642.68	emotional actions, states and processes general	18	144.78	the media: books
4	534.85	kin	19	128.57	drinks and alcohol
5	474.14	drama, the theatre and show business	20	123.24	general appearance and physical properties
6	411.73	happy	21	121.41	judgement of appearance: positive
7	395.35	personal names	22	118.81	the universe
8	344.98	interested/excited/energetic	23	117.75	sad
9	339.54	anatomy and physiology	24	115.02	moving, coming and going
10	301.38	like	25	105.74	existing
11	215.86	entire; maximum	26	103.26	food

12	211.97	the media: tv, radio and cinema	27	98.71	sports
13	207.41	getting and possession	28	97.91	time: period
14	181.78	entertainment generally	29	92.81	information technology and computing
15	180.70	personal relationship: general	30	91.96	evaluation: good

TABLE 9.6. TOP 30 POSITIVE KEY SEMANTIC DOMAINS IN THE BIPOLARDISORDER CORPUS COMPARED WITH THE 'MENTALILLNESS' CORPUS (LL CUT OFF 10.83, MIN. FREQ. 5, *P* < 0.001) CALCULATED USING WMATRIX (RAYSON, 2008)

The keywords shown in Table 9.5 show that some of the symptoms of bipolar disorder are being discussed; for example, 'manic', 'swings', 'moods', 'episodes', 'highs'. They also reveal a tendency to report issues to do with bipolar disorder in the context of celebrities with the condition or celebrity spokespeople such as Catherine Zeta-Jones and Stephen Fry, two public figures who have both openly discussed their experiences of bipolar disorder. The prevalence of names of celebrities in the BipolarDisorder corpus, and the relatively low number of keywords concerned with the symptoms of bipolar disorder, seems to suggest that the press are reporting on the particular experiences of celebrities instead of reporting on the symptoms of bipolar disorder more generally. This is problematic because, as the description of the symptoms above shows, the experience of bipolar disorder can differ enormously from person to person.

The key semantic domains offer some insight into the range of symptoms discussed in the corpus and into the depiction of bipolar disorder more generally. The high frequency of words and phrases concerning comparison (COMPARING: DIFFERENT) is explained by the fact that bipolar is tagged as belonging to this semantic category. In terms of the symptoms reported, the HAPPY, EMOTIONAL ACTIONS, STATES AND PROCESSES and the INTERESTED/EXCITED/ENERGETIC categories all contain words pertaining to the symptoms of bipolar disorder; for example, the majority of words contained in the EMOTIONAL ACTIONS, STATES AND PROCESSES category relate to the

phrase 'mood swings'. Moreover, the HAPPY category contains words pertaining to the symptoms of mania, e.g. 'highs', 'euphoric', 'over-elation' and the word 'mania' itself makes up the majority of the words coded by Wmatrix as 'interested/excited/energetic'. Further exploration of the ANATOMY AND PHYSIOLOGY category reveals 182 instances (348.52 pmw) of the lemma 'sleep'. These instances all occur in the context of lack of sleep, sleeplessness and fatigue which is in line with the symptoms of (hypo)mania. These findings suggest that newspaper representations of bipolar disorder do discuss some of the common symptoms of bipolar disorder; however, a concordance search reveals only 32 instances of 'hallucinations' and only 3 instances of 'hearing voices' in the whole BipolarDisorder corpus. This suggests that the newspaper reports do not routinely mention the full range of symptoms of bipolar disorder. This arguably does not fulfil the press' responsibility to inform the public. This is problematic when taken in the context of diagnosis statistics about bipolar disorder. Bipolar UK states that it takes an average of 10.5 years for a person with bipolar disorder to receive the correct diagnosis, with people getting an average of 3.5 misdiagnoses before this time (Bipolar UK, 2019). Furthermore, when we consider the stigma associated with voice hearing (Vilhauer, 2017), not mentioning this as a symptom of bipolar could arguably exacerbate stigma (i.e. by making it unmentionable).

Although the keyword list and the semantic domain analysis suggest that some symptoms of bipolar disorder are represented in the corpus, the most prototypical text of the BipolarDisorder corpus contains no description of the symptoms:

> ACTRESS Catherine Zeta-Jones has opened up about suffering from bipolar disorder. The 41-year-old said there was no need to suffer in silence and hoped publicity surrounding her treatment might help others. In a statement, the actress and mother to son Dylan, 10, and daughter Carys, eight, said: This is a disorder that affects millions of people and I am one of them. If my revelation of having bipolar II has encouraged one person to seek help, then it is worth it. There is no need to suffer silently and there is no shame in seeking help. Zeta Jones checked into rehab after helping husband Michael Douglas battle throat cancer. The Oscar-winning star spent five days getting treatment, reportedly at the Silver Hill psychiatric hospital in New Canaan, Connecticut, earlier this month. Her publicist, Cece Yorke, said at the time: After dealing with the stress of the past year, Catherine made the decision to check into a mental health facility for a brief stay to treat her bipolar II disorder. Bipolar disorder, formerly known as manic depression, is a severe mood disorder
>
> (*The Herald, Glasgow,* April 22, 2011)

The most prototypical text gives further evidence for the finding that mental illnesses (and specifically bipolar disorder) are often reported in the context of celebrity experiences of mental illness. In the article, no description of bipolar disorder or its symptoms is given apart from that is it a "severe mood disorder". This is disappointing, particularly as the article describes Zeta-Jones as having bipolar II, which means that there was an opportunity for the journalist to describe the difference between bipolar II and other types of bipolar disorder. Moreover, the article appears to suggest that Zeta-Jones' bipolar was caused by stress brought about by her husband's cancer diagnosis, and was treated in just 5 days. The experience described, then, does not accurately depict most people's experiences of bipolar disorder which is a long-term disorder that often requires people to manage their condition over their lifespan rather than seeking quick treatment.

Taken together, the keyness and prototypical text analysis of the BipolarDisorder corpus indicates that many of the symptoms of bipolar disorder are mentioned in the news reports on bipolar, however, many symptoms are left unreported, including the

psychotic symptoms of bipolar disorder including hallucinations and voice hearing. In order to fully represent the reality of bipolar disorder in the press, articles need to include more references to these symptoms in order to fully inform the public about the reality of bipolar disorder. Doing this will also help to destigmatise voice hearing and hallucinations by bringing discussion of these symptoms into the open. In addition to neglecting some of the symptoms of bipolar disorder, there are some other problematic reporting tendencies in reports on bipolar disorder such as using celebrity spokespeople or using celebrities as example cases of people with bipolar disorder, rather than discussing bipolar disorder in its societal context, i.e. as an illness that affects up to 2% of the UK population with a lifetime prevalence (Bipolar UK, 2019).

9.6. Obsessive compulsive disorders

In this section, I discuss the top keywords and key semantic domains in the OCD corpus in relation to the symptoms of obsessive disorders. Obsessive compulsive disorders include OCD and body dismorphia. OCD has two components. These are obsessions, which are repetitive intrusive thoughts or worries that cause anxiety, and compulsions, which are repetitive activities than people with OCD may do in order to temporarily ease the anxiety caused by obsessions, e.g. repeatedly checking a door is locked (*Mind*, 2019h). A common misconception about OCD is that it is to do with cleaning or being tidy. This is not the case. OCD is a condition that causes people who experience it to feel like they have no control over their thoughts and/or that not doing something in a particular way may cause bad things to happen. Many people with OCD report that they only have the obsessive side of OCD, which is sometimes referred to are 'Pure-O'. People who experience OCD may also experience other mental illnesses including depression and anxiety. There are also mental health problems that are related to OCD because they include repetitive thoughts or behaviours. These are perinatal OCD (the OCD experienced before or after birth),

body dysmorphic disorder, compulsive skin picking, trichotillomania (compulsive pulling out of hair), hoarding and obsessive compulsive personality disorder (*Mind*, 2019h).

The symptoms of OCD include fear of causing (or having already caused) harm, fear of contamination (or contaminating others), the ritualistic washing of hands or arranging objects and repeating words or phrases in your head. Tables 9.7 and 9.8 show the top 30 keywords and key semantic domains in the OCD corpus.

Rank	LL.	Item	Rank	LL	Item
1	32572.08	ocd	16	1038.751	my
2	7072.916	obsessive	17	1015.624	cleaning
3	6038.734	compulsive	18	953.842	wash
4	3196.534	rituals	19	889.967	tourette
5	2767.387	thoughts	20	837.235	behaviour
6	2180.545	washing	21	829.003	beckham
7	2130.468	obsessions	22	818.798	condition
8	2095.883	compulsions	23	808.37	ocds
9	2049.437	checking	24	790.263	instrusive
10	1730.725	germs	25	722.093	repetitive
11	1697.938	i	26	680.865	clean
12	1497.948	disorder	27	666.901	compulsion
13	1370.007	sufferers	28	666.306	ocdaction
14	1167.913	hands	29	663.204	check
15	1097.227	contamination	30	581.791	fear

TABLE 9.7. TOP 30 KEYWORDS IN THE OCD CORPUS COMPARED WITH ALL OTHER ILLNESS CORPORA MINUS THE 'MENTALILLNESS' CORPUS CALCULATED USING ANTCONC (ANTHONY, 2017).

Rank	LL.	Domain	Rank	LL	Domain
1	4748.38	PRONOUNS	16	321.60	UNWANTED
2	3343.08	CLEANING AND PERSONAL CARE	17	318.64	SENSORY: TOUCH
3	2663.82	INTERESTED/EXCITED/ENERGETIC	18	313.02	RELIGION AND THE SUPERNATURAL
4	1396.21	ANATOMY AND PHYSIOLOGY	19	269.06	THOUGHT, BELIEF
5	1012.28	OBJECTS GENERALLY	20	251.39	MOVING, COMING AND GOING

6	954.57	UNMATCHED	21	233.40	GETTING AND POSSESSION
7	839.51	LIVING CREATURES: ANIMALS, BIRDS, ETC.	22	231.27	GENERAL APPEARANCE AND PHYSICAL PROPERTIES
8	736.55	FEAR/SHOCK	23	217.76	WORRY
9	646.11	FREQUENT	24	204.27	COMPARING: USUAL
10	532.39	FOOD	25	198.76	MENTAL OBJECT: CONCEPTUAL OBJECT
11	477.85	INTERESTED/EXCITED/ENERGETIC	26	192.13	ELECTRICITY AND ELECTRICAL EQUIPMENT
12	457.59	CLOTHES AND PERSONAL BELONGINGS	27	187.81	PARTS OF BUILDINGS
13	404.47	LIKELY	28	184.36	LIKE
14	360.72	JUDGEMENT OF APPEARANCE: NEGATIVE	29	176.03	TEMPERATURE: COLD
15	331.48	JUDGEMENT OF APPEARANCE: POSITIVE	30	171.40	IF

TABLE 9.8. TOP 30 POSITIVE KEY SEMANTIC DOMAINS IN THE OCD CORPUS COMPARED WITH THE 'MENTALILLNESS' CORPUS (LL CUT OFF 10.83, MIN. FREQ. 5, *P* < 0.001) CALCULATED USING WMATRIX (RAYSON, 2008)

The top 30 keywords indicate that the two different parts of OCD (obsessions and compulsions) are represented in the corpus, as well as many words related to symptoms. These include 'rituals, 'thoughts', 'washing', 'checking', 'contamination', 'compulsion', 'check'. There are, however, some words that indicate an over-representation of certain facets of OCD such as contamination as is indicated by the keywords 'germs', 'washing', 'hands', 'cleaning', 'wash', 'clean'. In addition to this, as was the case with bipolar disorder, the OCD keywords also include reference to celebrities – for example, the footballer David Beckham – as the extract below shows:

> Medical experts say Beckham's weird rituals suggest he suffers from a version of obsessive compulsive disorder (OCD). This rare condition is caused by a chemical inbalance in the brain which makes victims repeatedly wash, clean or arrange their environment into "perfect" shapes. Beckham's bizarre obsessions emerged when he let cameras follow him for six months for a Christmas video for his fans.
>
> (*The People*, November 19, 2000)

As well as providing further evidence that celebrities are used as examples of people with mental illnesses, the extract shows that even in 2000, journalists were prepared to describe diagnosable medical illnesses as "weird" and "bizarre". Furthermore, the article incorrectly states that OCD is caused by chemical imbalances in the brain, which is not true – there is no one cause of OCD and causes differ from person to person. Again, this is an example of one person's (possible) experience of a mental illness (Beckham had not, at this point, announced that he had OCD) being presented as though it is the experience of all people experiencing OCD. Furthermore, the only symptoms mentioned are mostly compulsive and not obsessive, i.e. arranging and cleaning.

The over-representation of the compulsive side of OCD, i.e. washing and cleaning, is also visible in the key semantic domain list, where the semantic category CLEANING AND PERSONAL CARE is the second most statistically significant category. Similarly, the fourth and fifth most significant semantic categories ANATOMY and OBJECTS GENERALLY which includes words such as 'hands' (e.g. washing hands) and 'objects' (e.g. arranging objects) reiterates this finding. Moreover, the seventh most statistically significant category, LIVING CREATURES: ANIMALS, BIRDS, ETC., contains 297 instances of 'germs'.

Surprisingly, in the OCD corpus there is only 1 instance of 'intrusion*' and 146 instances of 'intrusive', compared with 418 instances of compulsion*; this is indicative of the skew towards the reporting of the compulsive component of OCD. The lack of

public understanding of OCD is evidenced in the most prototypical text in the OCD corpus (below), which is an advertisement for the charity OCD Action, which aims to raise awareness about OCD and recovery from OCD.

> The site is run by the specialist charity OCD Action with its committee comprising some of the UK's top specialists in obsessive compulsive disorder (OCD). It promotes recovery from obsessive compulsive disorders through understanding and awareness and gives information on the different types of OCDs . It also provides free information packs, contact details for support groups, discussion forums and organises OCD conferences.
>
> (*The Times,* April 4, 2006)

The fact that the most prototypical text does not include any symptoms, or description of the illness more generally, but instead refers readers to a place with that information, demonstrates that newspapers are printing ways that the public can be informed about OCD (i.e. through visiting third-party websites). However, descriptions of OCD that include the full range of symptoms – either from those who live with the condition or from medical professionals – are uncommon. In order for press to fully represent the reality of OCD, there is a need for more coverage of the obsessive side of OCD. Reporting only on the compulsion component of OCD, or skew in reporting towards this component, serves only to reiterate myths about OCD, such as that it is simply to do with germs and tidying.

9.7. Psychotic disorders

In this section, I discuss the top keywords and key semantic domains in the Psychosis corpus in relation to the symptoms of psychotic disorders. Psychotic disorders include psychosis and schizophrenia. For the reasons outlined in Section 9.3, I only discuss psychosis in this section as there is another subcorpus specifically for schizophrenia.

As a result, the psychosis corpus is not representative of the press coverage of schizophrenia. There are two main symptoms of psychosis. The first of these is hallucinations in which the person experiencing the psychotic episode will see things that others can't, sense things that are not actually happening or that have no cause (e.g. tasting something or feeling sensations on the skin or hearing voices). The voices a person may hear could be positive voices or disturbing voices that are hostile. (*Mind*, 2019i). The second main symptom of psychosis is delusions. Delusions can cause people to believe things, even when what they believe doesn't make sense logically. For example, a person may believe that that they are more powerful than they are, or that someone is trying to harm them (*Mind*, 2019i). These delusions are sometimes called delusions of grandeur and paranoid delusions, respectively. In addition to these symptoms, people experiencing psychosis may have disorganised thinking and speech resulting in them having racing thoughts or having flights of ideas where the link between one thing and another is not clear to their interlocutors. These symptoms are also symptoms of bipolar disorder and schizophrenia.

Tables 9.11 and 9.12 show the top 30 keywords and key semantic domains for the psychosis corpus.

Rank	LL.	Item	Rank	LL	Item
1	13439.108	cannabis	16	971.69	antipsychotics
2	11531.016	psychotic	17	937.904	breivik
3	9023.846	psychosis	18	861.603	patients
4	3650.694	dementia	19	813.572	users
5	3468.12	drugs	20	806.322	homes
6	2528.156	drug	21	753.872	medication
7	1623.453	skunk	22	712.812	hallucinations
8	1561.716	psychotics	23	632.462	smoking
9	1552.951	use	24	603.605	schizophrenia
10	1260.362	antipsychotic	25	566.277	nurofen
11	1205.911	mental	26	562.802	paranoid
12	1184.729	brady	27	561.157	risk
13	1166.143	anti	28	555.476	alzheimer
14	1074.09	care	29	532.663	psychiatric
15	1040.479	thc	30	529.009	rosenhan

TABLE 9.11. TOP 30 KEYWORDS IN THE PSYCHOSIS CORPUS COMPARED WITH ALL OTHER ILLNESS CORPORA MINUS THE 'MENTALILLNESS' CORPUS CALCULATED USING ANTCONC (ANTHONY, 2017)

Rank	LL.	Domain	Rank	LL	Domain
1	5208.12	SMOKING AND NON-MEDICAL DRUGS	16	100.99	LIVING CREATURES: ANIMALS, BIRDS, ETC.
2	1283.58	MEDICINES AND MEDICAL TREATMENT	17	100.32	TIME: OLD; GROWN-UP
3	1126.86	USING	18	76.68	PLANTS
4	479.00	ANATOMY AND PHYSIOLOGY	19	73.55	TOUGH/STRONG
5	422.20	PRONOUNS	20	68.33	LINEAR ORDER
6	352.13	SUBSTANCES AND MATERIALS GENERALLY	21	67.45	OPEN; FINDING; SHOWING
7	328.82	UNMATCHED	22	66.10	SPEED: FAST
8	165.49	CAUSE & EFFECT/CONNECTION	23	65.63	DRINKS AND ALCOHOL
9	163.42	DANGER	24	62.79	THE MEDIA: BOOKS
10	134.99	EVALUATION: FALSE	25	61.24	WEIGHT: HEAVY
11	119.84	SENSORY: SOUND	26	60.07	COMPARING: UNUSUAL
12	114.87	KIN	27	59.80	SPEECH: COMMUNICATIVE

13	109.06	TIME: OLD, NEW AND YOUNG; AGE	28	58.96	TIME: EARLY
14	106.00	OBJECTS GENERALLY	29	54.19	PERSONALITY TRAITS
15	102.09	TIME: NEW AND YOUNG	30	53.06	RELATIONSHIP: INTIMACY AND SEX

TABLE 9.12. TOP 30 POSITIVE KEY SEMANTIC DOMAINS IN THE PSYCHOSIS CORPUS COMPARED WITH THE 'MENTALILLNESS' CORPUS (LL CUT OFF 10.83, MIN. FREQ. 5, *P* < 0.001) CALCULATED USING WMATRIX (RAYSON, 2008)

What is most apparent from the top keywords and key semantic domains in the Psychosis corpus is the significant result for 'cannabis', 'skunk', 'drug', 'drugs' and 'smoking and non-medical drugs'. In fact, 'cannabis' appears in the Psychosis corpus 3,366.99 times per million words, which is more than the word 'psychosis' appears (2,016.29 pmw). On closer inspection, many of the instances that mention drugs in the corpus relate to drugs causing psychosis, for example, this extract from *The Times* in 2001 which reads "[I]n alarmist articles in The Times and elsewhere, they argued that scientific evidence shows that cannabis is addictive, causes personality change and psychosis". Drug taking, and particularly smoking cannabis, then, is a key theme in the Psychosis corpus. This is further evidenced by the third most key semantic category 'using' which includes phrases such as 'use of the drugs' and 'marijuana use'.

Drug taking for medicinal purposes is also a theme in the keyness analysis. For example the 'medicines and medical treatment' and 'substances and materials generally' categories, both of which refer to words and phrases concerning the treatment of psychosis. For example, 'antipsychotics', 'halperidol[81]' and 'molecules, chemicals' respectively.

What these findings show is that there is a tendency in the Psychosis corpus to report on the causes of psychosis rather than describing what the illness is and what the symptoms of it are. The only keyword related to descriptions or symptoms of psychosis are 'schizophrenia' and 'paranoid'. On further exploration, some of the key semantic domains do contain words and phrases concerned with the symptoms of

[81] Haloperidol is an anti-psychotic drug.

psychosis, however. These include the 'sensory: sound' and 'evaluation: false' categories which include 'hearing voices' and 'delusions' respectively. Despite these categories including these words, the frequencies of 'delusions' and 'hallucinations' are still comparatively low, with 'delusion*' occurring just 300.3 times pmw, and 'hallucin*' appearing just 314.6 times pmw. By comparison, the word 'Breivik' which relates to an isolated terrorist incident in which a man detonated a bomb and shot 69 people dead in Oslo, occurs 179.4 times pmw. What these findings indicate is that press reportage of psychosis and psychotic disorders over-represent links between psychosis and criminality, and under-represent information about psychosis and its symptoms.

The link between psychosis and criminality is also evidenced by the key semantic domain 'danger' which has not been present in any of the other keyness analyses of the illness subcorpora. Moreover, an analysis of the adjective predicates[82] of 'schizophrenia' (conducted using Word sketch tool on Sketch Engine) supports the finding that psychosis and criminality are linked as schizophrenia is the only illness in the corpus to collocate with the adjectives 'violent' and 'dangerous'. In addition, the most prototypical text for the Psychosis corpus is a news article reporting on a man with schizophrenia who fatally stabbed a woman:

> A SCHIZOPHRENIC stabbed a pensioner to death after his carers took him to buy alcohol. Martin Davies, who stabbed 66-year-old Gwen Poole at least four times with a breadknife, was yesterday ordered to be detained indefinitely. Cardiff Crown Court heard Davies, 23, obeyed voices in his head telling him to kill after he awoke from a nap, having earlier drunk the vodka he had bought with his carers
>
> (*South Wales Echo*, May 5, 2010)

[82] The Word sketch tool is based on collocations (using the LogDice statistical test).

The article does outline one of the symptoms of psychosis – hearing voices and hallucinations, but it does so in the context of a violent crime where the voices the person with schizophrenia heard were harmful commands. To return to the finding outlined in Section 9.5 (bipolar disorders), in which I argued that the psychotic symptoms of bipolar disorder are not mentioned and therefore become stigmatised due to being unmentionable, it appears that this is also true in the Psychosis corpus. The press neglect to describe these symptoms in detail outside of the fairly extreme cases in which, sadly, a person is hurt. Doing this creates links between mental illness (here specifically psychosis) and violent crime. This is at odds with research which has shown that any violent behavior in people with psychotic disorders such as schizophrenia is usually caused by substance abuse (*NHS*, 2019a), rather than as a direct cause of their diagnosis. Moreover, people with a mental illness are more likely to be the victim of a crime than the perpetrator (*Time to Change*, 2019b). Furthermore, it is interesting to note the absence of any celebrity spokespeople or celebrity examples in the keyword list, which further suggests that as a society psychotic disorders are not something that many people, celebrity or not, align with (if we are to take the topic matter discussed in newspaper articles as indicative of what is socially acceptable).

With this in mind, the representation of psychotic disorders in the press needs to focus more on descriptions of psychotic disorders and their symptoms, rather than on causes (e.g. drug use like cannabis) or isolated criminal incidents such as the case described in the most prototypical text and the case of Anders Breivik. Focusing on violent crime as well as substance abuse does not represent the many people with psychotic disorders who are living well and managing their condition. Furthermore, this type of press coverage actively stigmatises psychotic disorders because these articles suggest that psychosis is self-inflicted (i.e. caused by drug misuse). This claim is supported by the findings of research conducted by Mann & Himlein (2004), who discovered that schizophrenia was more stigmatised that other illnesses, such as depression. Similarly, research conducted by Nawka et al. (2012) in the Czech Republic found that news articles reporting psychotic disorders were more likely to

contain descriptions of aggressive behaviour and homicide. Furthermore, Goulden et al. (2011: 5) found that in contrast to other mental illnesses, schizophrenia was more likely to be mentioned in the context of 'bad news'. Overall, what my findings combined with exiting research suggest is that the press do not represent psychotic disorders accurately because these disorder types are taboo.

9.8. Anxiety disorders

In this section, I discuss the top keywords and key semantic domains in the Anxiety corpus in relation to the symptoms of anxiety disorders. Anxiety disorders include a range of illnesses including social anxiety disorder, generalised anxiety disorder and body dysmorphia. Anxiety disorders also include phobias. Phobias are classed as anxiety disorders when the phobia lasts more than six months, has a serious impact on the everyday life on the person, or when the phobia is classed as extreme given the realistic risk of the phobia (*Mind*, 2019j). Phobias are often subcategorised as specific and complex phobias. The first of these phobia types is concerned with phobias of specific objects or situations such as heights or particular animals, whereas the latter refers to phobias that are more disruptive than the specific phobias, and that often affect the everyday life of the person experiencing the phobia. The most common complex phobias are social phobia and agoraphobia. Social phobia, which is sometimes called social anxiety disorder causes those who experience it to feel high levels of anxiety when in social situations such as public speaking or eating in front others. As a result, people who experience social anxiety may avoid social situations, which can inhibit their ability to work and perform everyday tasks (*Mind*, 2019j). Similarly, the symptoms of agoraphobia include high levels of anxiety when in certain places or situations. As a result, people with agoraphobia may avoid leaving their home or being in a crowd of people (*Mind*, 2019j). Symptoms of phobias include

feeling dizzy or faint, palpitations, nausea and numbness. In some situations, these symptoms can trigger a panic attack (*Mind*, 2019j)

Tables 9.13 and 9.14 show the top 30 keywords and key semantic domains for the Anxiety corpus.

Rank	LL.	Item	Rank	LL	Item
1	40908.881	anxiety	16	1415.95	breathing
2	5231.574	your	17	1391.817	president
3	4949.986	lewinsky	18	1324.126	can
4	4685.534	fear	19	1169.712	exam
5	3893.807	panic	20	1029.225	yoga
6	3672.514	you	21	893.828	hypnotherapy
7	3518.29	anxious	22	890.387	techniques
8	3475.994	stress	23	831.016	dentist
9	2710.694	phobia	24	824.333	currie
10	2115.377	or	25	812.234	dental
11	1822.839	attacks	26	811.224	relax
12	1631.003	worry	27	810.148	fears
13	1609.089	sleep	28	783.768	anxieties
14	1544.352	phobias	29	773.887	botton
15	1525.891	relaxation	30	754.994	attack

TABLE 9.13. TOP 30 KEYWORDS IN THE ANXIETY CORPUS COMPARED WITH ALL OTHER ILLNESS CORPORA MINUS THE 'MENTALILLNESS' CORPUS CALCULATED USING ANTCONC (ANTHONY, 2017)

Rank	LL.	Domain	Rank	LL	Domain
1	6705.37	WORRY	16	484.97	EDUCATION IN GENERAL
2	6387.89	ANATOMY AND PHYSIOLOGY	17	391.75	DRINKS AND ALCOHOL
3	2631.45	PRONOUNS	18	388.66	MEASUREMENT: LENGTH & HEIGHT
4	2534.82	FOOD	19	380.46	COMPARING: USUAL
5	1796.12	FEAR/SHOCK	20	371.57	EMOTIONAL ACTIONS, STATES AND PROCESSES GENERAL
6	1146.76	SPORTS	21	356.66	PLANTS
7	976.85	CALM	22	351.68	VEHICLES AND TRANSPORT ON LAND

8	798.36	UNMATCHED	23	336.89	MEASUREMENT: WEIGHT
9	708.37	LIVING CREATURES: ANIMALS, BIRDS, ETC.	24	335.46	LIKELY
10	653.33	OBJECTS GENERALLY	25	309.57	TIME: MOMENTARY
11	643.70	SUBSTANCES AND MATERIALS GENERALLY	26	284.90	TIME: NEW AND YOUNG
12	618.45	SUBSTANCES AND MATERIALS: LIQUID	27	280.45	MOVING, COMING AND GOING
13	592.88	FLYING AND AIRCRAFT	28	264.30	IF
14	557.97	RELATIONSHIP: INTIMACY AND SEX	29	236.79	SHAPE
15	538.79	THOUGHT, BELIEF	30	232.09	ELECTRICITY AND ELECTRICAL EQUIPMENT

TABLE 9.14. TOP 30 POSITIVE KEY SEMANTIC DOMAINS IN THE ANXIETY CORPUS COMPARED WITH THE 'MENTALILLNESS' CORPUS (LL CUT OFF 10.83, MIN. FREQ. 5, *P* < 0.001) CALCULATED USING WMATRIX (RAYSON, 2008)

Many of the top keywords and top key semantic domains (in particular, the 'worry' and 'fear/shock' categories) include many of the symptoms of anxiety disorders such as 'anxiety', 'stress', 'distress' and 'fear', 'panic', 'terror', 'dread' respectively. Further to this, the semantic key domain 'flying and aircraft' contains instances in which a fear of flying is being described. A collocation analysis of the phrase "a fear of" reveals that a fear of flying is the most discussed fear in the corpus. This is shown in the top 5 collocates of "fear of" which are 'flying', 'spiders', 'spaces', 'heights', 'failure'. Further to this, the top 10 collocates (L3, R3) of 'phobia' are 'social', 'school', 'develop', 'specific', 'attacks', 'School', 'fears', 'agoraphobia', 'panic' and 'dental'. What this shows is that the press do discuss the difference between general and specific phobias, and do discuss the two most common complex phobias: social phobia and agoraphobia. Furthermore, the collocates of 'phobia' indicate that the press accurately describe the process of getting a phobia, via 'developing'. Closer concordance analysis of this lexical item shows that many articles describe the causes and triggers of phobias, i.e. they may develop from a trauma or worry around having a panic attack in response to something.

Despite the fact that the articles reporting on phobias do seem to use the correct terminology to refer to the different phobia types as well as describing the causes of the phobias accurately, keyness analysis does not suggest that the press reports fully describe the symptoms experienced by people with complex phobias. Aside from 'panic attacks', only general symptoms are represented in the keyword list (e.g. anxiety and stress). The lack of representation of symptoms could be due to the fact that many of the symptoms of anxiety disorders are also symptoms of other mental illnesses, which may mean that words related to these symptoms would not come out as key in the keyness analysis (because the MentalIllness corpus also contains them). This may be a contributing factor to the absence of many of the symptoms of the complex phobias in the corpus, however the most prototypical text in the corpus also does not include any mention of symptoms:

> ANXIETY Care is soon to start a new volunteer training course in Redbridge. Successful trainees will help people with anxiety disorders to return to normal lives through home visiting and group work. For details telephone 8262 8891/2 or write to the training officer, Anxiety Care, 19 Mansfield Road, Ilford or email anxietycare@aol.com
>
> (*This is Local London*, May 23, 2001)

What these findings appear to indicate is that the complex phobias are underrepresented as illnesses on their own as the most well-represented phobia in the corpus is concerned with fear of flying which is a specific phobia. In order to better describe the reality of anxiety disorders, the press would do well to provide more information about the full range of anxiety disorders, including the complex phobias to inform the public that complex phobias are illnesses in themselves and not simply symptoms of other illnesses.

9.9. Eating disorders

In this section, I discuss the top keywords and key semantic domains in the EatingDisorder corpus in relation to the symptoms of eating disorders. Eating disorders include anorexia, bulimia and binge-eating disorder. The general symptoms of eating disorders include worrying about your weight, eating too little or too much food, having strict routines around food and exercising too much.

Anorexia (also called anorexia nervosa) is a condition where a person deliberately does not eat enough food for their body which results in them having a body weight that is under what is healthy. Anorexia is often thought to be linked to losing weight, but the reasons for anorexia are often closely linked to low self-esteem and issues with self-image. Symptoms of anorexia include missing meals, having rituals around eating and feeling like you are overweight. Indirect symptoms of anorexia are amenorrhea[83] in women, headaches and problems sleeping (NHS, 2019b).

Bulimia (also called bulimia nervosa) is an eating disorder characterised by eating large quantities of food over a short period of time and then purposely making yourself vomit because you feel ashamed and guilty. This process is often called binging (from binge-eating) and purging (vomiting). People with bulimia may also take laxatives or exercise excessively. Symptoms of bulimia include eating in secret, feeling ashamed, feeling that you are overweight and feeling depressed.

Binge-eating disorder (sometimes called compulsive eating) is an illness that causes people with it to eat very large quantities of food in short periods of time resulting in the person feeling uncomfortably full. Symptoms of binge-eating disorder include feeling out of control of your eating, eating in secret and feeling ashamed after binging.

Tables 9.15 and 9.16 show the top 30 keywords and key semantic domains for the EatingDisorder corpus.

[83] *Amenorrhea* is the medical term for the absence of menstruation.

Rank	LL.	Item	Rank	LL	Item
1	97288.006	eating	16	9100.503	fat
2	57761.024	anorexia	17	8639.934	size
3	36296.462	disorders	18	8289.416	fashion
4	25597.46	she	19	8059.952	nervosa
5	24944.929	weight	20	6029.912	anorexics
6	23545.314	bulimia	21	5914.776	skinny
7	18429.889	her	22	5721.398	stone
8	15989.135	i	23	4931.9	diet
9	14754.601	food	24	4709.521	my
10	14367.483	thin	25	4356.182	dieting
11	14366.238	anorexia	26	3705.346	diana
12	13172.097	eat	27	3637.218	bulimic
13	12136.13	models	28	3629.804	binge
14	12103.046	girls	29	3573.845	women
15	10190.163	disorder	30	3408.518	calories

TABLE 9.15. TOP 30 KEYWORDS IN THE EATINGDISORDER CORPUS COMPARED WITH ALL OTHER ILLNESS CORPORA MINUS THE 'MENTALILLNESS' CORPUS CALCULATED USING ANTCONC (ANTHONY, 2017)

Rank	LL.	Domain	Rank	LL	Domain
1	31092.70	FOOD	16	917.51	JUDGEMENT OF APPEARANCE: POSITIVE
2	13106.27	PRONOUNS	17	868.82	MEASUREMENT: AREA
3	7653.10	MEASUREMENTS: WEIGHT	18	748.00	SUBSTANCES AND MATERIALS GENERALLY
4	5087.71	ANATOMY AND PHYSIOLOGY	19	687.10	EXISTING
5	4828.87	PEOPLE: FEMALE	20	570.55	EXCESSIVE DRINKING
6	1790.67	KIN	21	560.15	SEEM
7	1782.16	SHORT AND NARROW	22	559.50	RELATIONSHIP: INTIMACY AND SEX
8	1745.10	CLOTHES AND PERSONAL BELONGINGS	23	558.56	EXCLUSIVERIZERS/PARTICULARS
9	1410.48	MEASUREMENT: SIZE	24	554.59	SHAPE
10	1382.41	TIME: OLD NEW AND YOUNG; AGE	25	539.13	DEGREE: BOOSTERS

11	1371.42	LACK OF FOOD	26	518.62	LIKE
12	1262.45	OBJECTS GENERALLY	27	510.28	THOUGHT, BELIEF
13	1107.14	WEIGHT: LIGHT	28	456.19	TIME: NEW AND YOUNG
14	1028.16	SPORTS	29	419.50	NEGATIVE
15	1005.46	MEASUREMENT	30	418.92	HEALTHY

TABLE 9.16. TOP 30 POSITIVE KEY SEMANTIC DOMAINS IN THE EATINGDISORDER CORPUS COMPARED WITH THE 'MENTALILLNESS' CORPUS (LL CUT OFF 10.83, MIN. FREQ. 5, *P* < 0.001) CALCULATED USING WMATRIX (RAYSON, 2008)

The top 30 keywords indicate that there is a clear tendency in the corpus to talk about eating disorders in the context of weight, particularly using evaluative terms about weight such as 'skinny', 'thin' and 'fat'. Moreover, the top keywords include references to weight loss through 'dieting' and 'calories' and 'stones'. This focus on weight loss in the depiction of eating disorders is unhelpful because, as stated in the description of the eating disorders above, the causes of eating disorders are often much more to do with the self-esteem of the person rather than weight loss. Attributing eating disorders to a desire to lose weight is far more simplistic than the reality of the illnesses which often occur with other types of self-harm.

A further indication that the representation of eating disorders is unrealistic is the focus on women in the keywords, indicated by 'she', 'her', 'girls' and 'women'. This finding is also supported by an analysis of the top adjective predicates[84] of 'eating disorder' (conducted using the Word sketch function on Sketch Engine) which shows that 'male' is an adjective collocate of 'eating disorder'. This finding suggests that in some cases, 'eating disorder' is marked i.e. with 'male' to convey that the person with an eating disorder is male. This is indicative that the 'eating disorder' itself is, in some cases, presumed to be a female illness. Whilst eating disorders are more prevalent in women and girls, this apparent tendency to present eating disorders a female illness in the corpus is incorrect. Latest statistics from Anorexia and Bulimia Care state that around a quarter of all people with an eating disorder are male (Anorexia & Bulimia

[84] The Word sketch tool is based on collocations (using the LogDice statistical test).

Care, 2019), with diagnoses of eating disorders rising by 27% in males (Micali et al., 2013).

In keeping with the general trend in newspaper reports on mental illness, the EatingDisorder corpus also includes celebrity examples. These include the keywords 'models' and 'diana', the latter relating to Princess Diana who reportedly had bulimia. The themes identified in the keyword list are also supported by the key semantic domains such as MEASUREMENT: WEIGHT, MEASUREMENT: SIZE, JUDGEMENT OF APPEARANCE: POSITIVE and WEIGHT: LIGHT. Closer inspection of these key semantic domains support the finding the press that over-represent size and weight in their description of eating disorders; for example, the MEASUREMENT: SIZE semantic category in which the majority of words pertain to clothes sizes, such as size 10, size 8, etc. What is also apparent from closer inspection of these categories is that many of the symptoms of eating disorders are not present in the articles. In fact, the EatingDisorder corpus has the lowest number of instances of the word 'symptom*' across all the illness subcorpora occurring just 176.33 times pmw (the PTSD corpus had the highest with 839.37 instances pmw). Concordance analysis of 'symptoms' reveals that when symptoms are discussed, in many cases they are discussed thoroughly and realistically, often mentioning eating disorders in men. Moreover, a search for 'self-esteem' reveals that there are 925 instances of this phrase (173.33 instances pmw) in the corpus which is positive; however, arguably this phrase ought to appear much more frequently given that self-esteem issues are at the root of many eating disorders. The frequency of 'self-esteem' in the corpus translates to around 0.02% of all tokens in the corpus.

It is the case, however, that articles that do discuss symptoms appear to be lower in frequency, as is indicated by keyness analysis. Further exploration of symptoms in the corpus reveal that the phrase 'feel* ashamed' (which includes 'felt', 'feel' and 'feeling') occurs just 35 times in the corpus (6.56 instances pmw) and 'menstrua*' occurs 90 times (16.86 instances pmw). In may be expected from these findings, the most prototypical text of the EatingDisorder corpus features no symptoms:

> TWO-thirds of people with eating disorders risk their lives by waiting more than six months for help. Sixty-two per cent of anorexics and bulimics crucially delay getting care, says a report by the Eating Disorders Association out tomorrow. Sufferers - estimated at 1.1million - are still often misunderstood and misdiagnosed. Nearly half blame the emphasis on skinny models for their disorders. An EDA spokeswoman said: "We're calling on parents, doctors and the media to hear what young people have to say about eating disorders."
>
> (*Sunday Mirror*, February 5, 2006)

Overall, the EatingDisorder corpus indicates that the press representations of EatingDisorders do not accurately describe the reality of eating disorders or their symptoms. Keyness analysis indicates that the topic of many of the articles contained in the corpus is weight loss and physical appearance, particularly weight loss in women. This is evidenced by the female pronouns in the corpus and the reference to women's clothes sizes in the key semantic domain analysis. For the newspapers to accurately represent eating disorders, more space needs to be dedicated to describing the nature of eating disorders – i.e. the fact that they are not simply about losing weight – as well as more thorough description of the symptoms and people eating disorders affect. By doing this, the public will be better informed about eating disorders and how to spot the early signs of these in themselves and others.

9.10. Depressive disorders

Depressive disorders include depression, seasonal affective disorder (SAD) and postnatal depression (where depression occurs after birth). Depressive disorders are characterised by low mood, feeling disengaged from activities you used to find pleasure in and feeling hopeless. Unlike periods of sadness, which are natural in everybody, depressive disorders are those that last a long time and which are

disabling. The symptoms of depression include feeling numb, feeling suicidal and like you want to harm yourself, feeling isolated, being overtired and losing interest in sex. In some cases of severe depression, the person experiencing the depression can also experience psychotic symptoms such as delusions and hallucinations (*Mind*, 2019k).

Seasonal affective disorder (SAD) is a type of depression that people only experience at particular times of year, or because of particular types of weather. People who experience SAD experience the same symptoms of depression but only during certain times of the year.

Postnatal depression (sometimes referred to as PND) is the depression encountered after giving birth. The symptoms of PND are very similar to depression but include additional symptoms such as feeling hostile towards your baby or your partner, or feeling like you are struggling to bond with your baby.

Tables 9.17 and 9.18 show the top 30 keywords and key semantic domains for the Depression corpus.

Rank	LL.	Item	Rank	LL	Item
1	58891.844	depression	16	2685.97	natal
2	17524.88	depressed	17	2568.651	depressing
3	6313.756	suicide	18	2567.856	english
4	5036.374	depressants	19	2476.602	print
5	4787.859	p	20	2424.387	i
6	4664.23	antidepressants	21	2325.922	ltd
7	4199.802	prozac	22	2315.615	mood
8	3788.487	his	23	2256.595	postnatal
9	3416.851	copyright	24	2247.397	inquest
10	3348.065	you	25	2149.647	independent
11	3134.292	anti	26	2117.559	seroxat
12	3107.112	documents	27	2116.734	depressant
13	2943.179	depressive	28	2106.371	has
14	2800.013	pounds	29	2096.045	market
15	2762.89	he	30	1858.602	suicidal

TABLE 9.17. TOP 30 KEYWORDS IN THE DEPRESSION CORPUS COMPARED WITH ALL OTHER ILLNESS CORPORA MINUS THE 'MENTALILLNESS' CORPUS CALCULATED USING ANTCONC (ANTHONY, 2017)

Rank	LL.	Domain	Rank	LL	Domain
1	9691.84	UNMATCHED	16	747.13	SUBSTANCES AND MATERIALS GENERALLY
2	2308.04	FOOD	17	730.98	SAD
3	2268.69	ANATOMY AND PHYSIOLOGY	18	707.99	JUDGEMENT OF APPEARANCE: POSITIVE
4	1674.85	BUSINESS: SELLING	19	678.88	MUSIC AND RELATED ACTIVITIES
5	1619.59	BUSINESS: GENERALLY	20	668.53	WEATHER
6	1613.11	PERSONAL NAMES	21	665.99	NUMBERS
7	1471.22	GEOGRAPHICAL NAMES	22	653.91	ENTERTAINMENT GENERALLY
8	1324.19	MONEY GENERALLY	23	626.65	SHAPE
9	1089.71	COLOUR AND COLOUR PATTERNS	24	598.97	CLOTHES AND PERSONAL BELONGINGS
10	1015.57	THE MEDIA: BOOKS	25	576.16	DRINKS AND ALCOHOL
11	940.31	SUBSTANCES AND MATERIALS: LIQUID	26	572.49	SUBSTANCES AND MATERIALS: SOLID
12	926.53	PAPER DOCUMENTS AND WRITING	27	568.46	PEOPLE: MALE
13	857.21	RELATIONSHIP: INTIMACY AND SEX	28	558.28	RELIGION AND THE SUPERNATURAL
14	778.57	LIVING CREATURES: ANIMALS, BIRDS, ETC.	29	548.15	TIME: PERIOD
15	766.40	GEOGRAPHICAL TERMS	30	540.54	LANGUAGE, SPEECH AND GRAMMAR

TABLE 9.18. TOP 30 POSITIVE KEY SEMANTIC DOMAINS IN THE DEPRESSION CORPUS COMPARED WITH THE 'MENTALILLNESS' CORPUS (LL CUT OFF 10.83, MIN. FREQ. 5, $P < 0.001$) CALCULATED USING WMATRIX (RAYSON, 2008)

The keyword list indicates that some of the symptoms of depression are represented in the news coverage. For example, 'suicide', 'suicidal, mood'. Furthermore, the key semantic domain 'food' also reveals that eating disorders are represented in the Depression corpus which is in line with the fact that many people with an eating disorder also experience depression. In addition, a closer inspection of the 'relationship: intimacy and sex' category reveals that the corpus contains words that

pertain to the reduced libido in people with depression. Furthermore, collocates of 'sex' in the corpus include 'drive' (e.g. "reduced sex drive") and 'interest' (e.g. "loss of interest in sex"). Furthermore, 'appetite' is also a collocate of 'sex' which, on closer inspection, is used in reference to changes (particularly loss) of appetite in a list of symptoms of depression. An example of this is shown below:

> Clinical depression is not a fancy name for the blues - it's a deeply debilitating biological illness affecting some 15 per cent of people some time in their lives. Sleep patterns, appetite and **sex** drive are all adversely affected. Sufferers feel worthless, are indecisive and unable to concentrate. And the condition is not confined to bored suburban housewives. Winston Churchill suffered terribly from "the black dog of depression" and actor Jim Carrey took Prozac to combat depressive bouts
>
> (*Daily Mirror*, April 29, 1996)

As well as showing a list of symptoms, the extract above also provides another example of where celebrities are used as examples of people who have a mental illness. This is also a feature of the most prototypical text in the Depression corpus which is about the boxer Ricky Hatton's experience of depression:

> RICKY HATTON has revealed just how close he came to committing suicide during his three-year retirement from boxing. The 34-year-old said his girlfriend often had to prise a knife from his hand as depression took its toll on his life. Hatton, who has had well-publicised battles with drink, drugs and depression since May 2009 after his loss to Manny Pacquiao in his last fight, will return to the ring against Ukraine's Vyacheslav Senchenko next month. He said: "I was near to a nervous breakdown, depression, suicidal. Most mornings my girlfriend would have to come downstairs and take a knife out of my hand. I had a knife at my wrists, I was in a really bad way, just hysterically crying for no reason. "I've always liked a little bit of a drink, but my drinking had gone way off the Richter scale, I was having blackouts. "And even if I was stonecold sober I was trying to kill myself.
>
> (*Daily Post*, October 29, 2012)

9.11. Exploring symptoms through Sketch Thesaurus

So far in this chapter I have based my analysis of how accurately the symptoms of particular illnesses are represented in the corpus on keyness analyses and prototypical text analysis. Such analyses are useful in determining what is idiosyncratic about each illness subcorpus, and seeing how each illness is reported on. However, in this section of the chapter, I explore the collocations of each disorder type. The reason for doing this is to explore whether collocation analysis reveals any symptoms or overlaps in the reporting of illnesses that were not apparent from keyness analysis. In order to carry out the collocation analysis, I conducted the final two steps in Process 2. These are 'use concordance analysis using syntactic search frame [query term and…] to see what symptoms, or other mental illnesses the query term is being textually equated with' and 'conduct Sketch Thesaurus search to compare findings'.

Table 9.19 shows the collocates that pertain to symptoms or other mental illnesses of [query term and…]. Collocates that co-occur with more than one illness are colour-coded to show overlap between the illness types.

PTSD and	Dissociation and	OCD and	Psychosis and	Agoraphobia and	Anorexia and	Depression and
schizophrenia	splitting	phobia	schizophrenia	claustrophobia	bulimia	anxiety
depression	self-disgust	ADHD	paranoia	social phobia	obesity	stress
stress	alienation	panic attacks	hallucinations	panic attacks	alcoholism	schizophrenia
anxiety	detachment	depression	neurosis	panic disorder	eating disorder	suicide
	claustrophobia	anorexia	Personality disorder	insomnia	depression	alcoholism
	trauma		delusions	anxiety	self-harm	Panic attacks
	disturbance		episodes	depression	addiction	insomnia
	delusion		depression	stress	self-mutilation	mood swings
	abuse		schizoaffective psychosis			manic depression
	depression					

TABLE 9.19. COLLOCATES OF QUERY TERM + AND IN THE MI 1984-2014 CORPUS (LOGDICE STATISTICAL CALCULATION, RANKED BY LOGDICE SCORE).

Table 9.19 reveals some symptoms and links between illnesses that were not revealed in the keyness analysis. For example, the collocation analysis of 'dissociation and…' is much more revealing of the symptoms of dissociative disorders than the keyness analysis indicated. This suggests that, despite the corpus being very small, the symptoms of dissociation are represented, for example, 'detachment' (1 instance), and 'delusions' (1 instance). Moreover, the causes of dissociative disorders such as 'trauma' (5 instances), and 'child abuse' (1 instance) are present. In addition, the collocation analysis also reveals some of the symptoms of other disorders that were not revealed in the keyness analysis. These include the collocate 'claustrophobia' of 'agoraphobia and…' (17 instances) and 'insomnia' for 'depression and…' (224 instances). Furthermore, the collocates shown in Table 9.19 reveal links between symptoms and illnesses that we may not associate such as 'alcoholism' and 'anorexia' (23 instances). Alcoholism is not listed as a linked illness or a symptom of eating disorders on the *Mind* website, however research shows that alcohol and eating disorders frequently co-occur (Grilo et al., 2002). Moreover, the keyness analysis of the EatingDisorder corpus did not reveal any mention of eating disorders being concerned with low self-esteem, however the collocation analysis reveals that 'self harm' (18 instances) is a collocate of 'anorexia and…', which shows that there is a link in the corpus between low self-esteem and eating disorders. This is due to the fact that self-harm is a sign of low self-esteem (*NHS*, 2019c). Table 9.19, then, has revealed symptoms previously uncovered by keyness analysis, but has also revealed links between illnesses and symptoms that even I as an analyst was unaware of. However, the frequencies of the collocates mentioned in this section are very few which means that whilst these symptoms may be present, they are not common.

In addition to revealing links between symptoms and illnesses, Table 9.19 also reveals links between the illness types. For example, people who experience PTSD also experience anxiety and depression, both of which are

collocates of 'PTSD and…'. Further, people who experience dissociative disorders and psychotic disorders may experience depression and 'depression' collocates with 'dissociation and…' and 'psychosis and…'. Similarly, links between OCD and eating disorders can be observed in Table 9.19, where 'OCD and…' collocates with 'anorexia'. This reflects the reality of eating disorders which are often linked with obsessive or compulsive behaviours. What these collocations indicate, then, is that similar, or linked illnesses are being discussed in the press together, which is a positive thing when the purpose of the press is to inform the public about mental illnesses. What is also interesting to note about the collocates shown in Table 9.19 is that 'depression' is the only illness to collocate with all the illnesses listed. This is indicative that depression is a commonly reported illness and may be the most societally acceptable illness as a result of that fact. The fact that depression is the most commonly reported illness is interesting when taken in the context of previous research which has shown that depictions of depression in the press have become more positive and less stigmatised over time (Goulden et al., 2011), whilst we cannot be sure whether the direction of this change over time was borne from greater exposure in the press, or societal acceptance which in turn resulted in a greater number of reports on depression , this finding is interesting because it suggests that the less stigmatised the illness, the more it is reported on. It may be the case then, that exposure of lesser reported illnesses (like, for example, dissociative identity disorder) may result in less stigma around them.

Table 9.19 shows the collocations of words in a fairly specific syntactic frame; however, these collocational patterns are also observable more broadly in the Sketch Thesaurus visualisation for each illness type which provides insight into how the illnesses are linked in the corpus. The rationale for using the Sketch Thesaurus tool for each illness type is that the Sketch Thesaurus tool generates usage-based synonyms, i.e. closely related words based on the surrounding context of those words (based on collocation). This means that

words that are closely related must share similar linguistic contexts. Figure 9.2 shows words that are usage-based synonyms of a query word (the word in the centre of the circle) based on how many collocates the two words share. The concentric circles relate to the statistical significance of the collocation (the higher the score, the more significant the collocation and the closer to the synonym is to query word). The size of the circle for each word relates to how many other words that word also collocates with. For example, the query term 'anorexia' collocates with 'bulimia' more statistically significantly than it does with 'depression', and 'bulimia' is a collocate of fewer words than 'depression' is. Sketch Thesaurus visualisations, then, are indicative of how closely related different illness types are in the corpus, and as such provides another way of visualizing the links between illnesses.

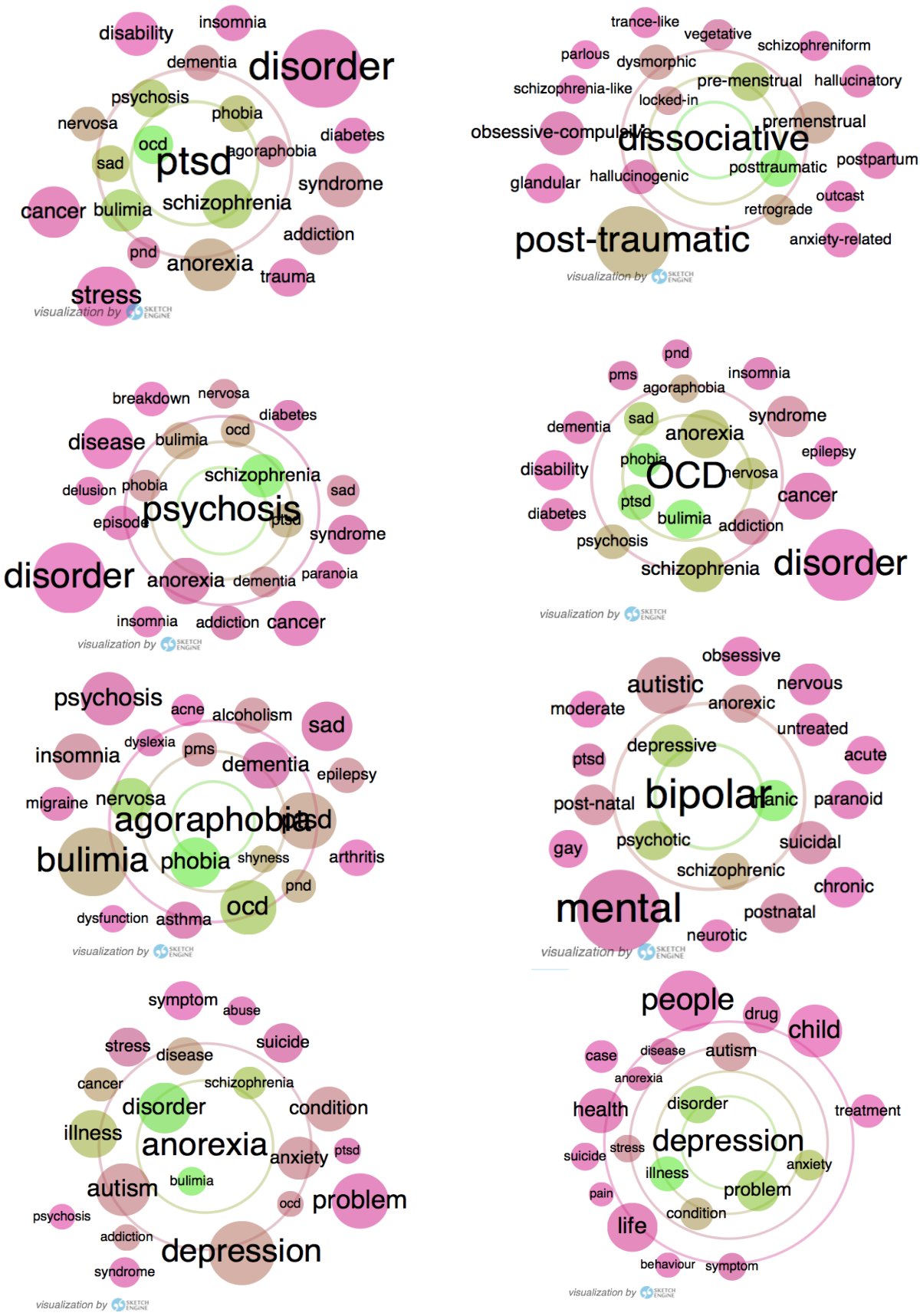


FIGURE 9.2. SKETCH THESAURUS VISUALISATIONS FOR EACH ILLNESS TYPE (CALCULATED USING SKETCH ENGINE).

9.12. Conclusion

In this chapter I have explored whether the symptoms of each disorder type are accurately represented in the new reports contained in the corpus. I have showed that corpus linguistic techniques, particularly keyness and collocation analysis provide the means to target links between mental illnesses and symptoms that may not be interpretatively significant to researchers using qualitative analyses only (for example, the link between alcoholism and anorexia). I have showed how a mixed-methods approach, i.e. keyness and collocation analysis combined with qualitative analysis of prototypical texts) provides a means of cross-comparison between quantitative and qualitative interpretation. Furthermore, I showed in Section 9.1 how the social context of a societal issue such as mental illness can be incorporated into linguistic analysis (through the steps I describe in Table 9.1).

In combining qualitative and quantitative methods from linguistics, I have built on previous research into the representation of symptoms in news reports on mental illness. Such as that reported in the introduction to this chapter conducted by Wahl et al. (2002). Furthermore, I have showed that the computational linguistic analysis of large datasets combined with qualitative analysis is much more revealing of the representation of specific mental illnesses than the method adopted in research like that conducted by Wahl et al. (2002). The reason for this is that in comparison to Wahl et al. (2002), who were only able to report that news articles rarely reported the symptoms of mental illness, the research I conducted here shows *what* symptoms were present in news reports and gives systematic and robust linguistic evidence for the how those symptoms were reported. Moreover, in contrast to Wahl et al. (2002), I have showed what symptoms occur in relation to specific illnesses. Taken together, the methods I have used in this corpus have resulted in more specific and replicable findings. Furthermore, keyness analysis has provided

insight into the general themes in mental illness reporting such as the tendency to report on celebrities with mental illnesses. The identification of this theme in some (but not all) of the illness subcorpora provides interesting new research avenues worthy of exploration such as why it is the case that celebrity names were statistically significant in the analysis of news reports on bipolar disorder and depression but not in the analysis of psychotic disorders like schizophrenia. It could be argued (as I allude to in Section 9.7 'Psychotic Disorders') that the absence of celebrity names associated with psychotic disorders is indicative that psychotic disorders are more stigmatised in the press than other illnesses, because psychotic disorders are not something that people (celebrity or not) align with. The theme of referring to the names of celebrities in reference to specific illnesses, then, is one way that linguistic analysis could potentially give insight into levels of societal stigma between illness types (e.g. through an analysis of what illnesses are linked with celebrity names).

In the next chapter, I revisit the research questions and findings reported in the analysis chapters and conclude this thesis.

10. Conclusion

In this thesis I have explored the representation of mental illness in the UK press. The reason for doing this is that the majority of research into press representations of mental illness to date have been conducted in the field of psychiatry, and as a result have not studied the language used to discuss mental illness as an object of study in its own right. This has resulted in the previous research being focused on the analysis of predetermined themes such as the depiction of people with mental illness as violent and criminal (e.g. Schomerus et al., 2012), and relatedly, on identifying stigma in news reports. Although some of the studies in psychiatry do note that language plays a role in the creation of stigma (e.g. Rhydderch et al., 2016), language in those studies is only mentioned as tangential to the analysis being conducted. Furthermore, as I discussed in the introduction to this thesis, it is increasingly the case that anti-stigma initiatives are promoting prescribed linguistic forms for discussing mental illness in the news and yet there exists no comprehensive account of the linguistic features of news reports on mental illness generally. In this thesis I have addressed this research gap. I have done this through designing and constructing the first and largest[85] corpus that contains UK local and national news articles on mental illness – the MI 1984-2014 corpus. Moreover, the design of the corpus means that diachronic and synchronic analyses can be conducted systematically within one dataset.

Combined with systematic linguistic analysis, the size of the corpus and the scope for different analysis types afforded by it means that the research I report in this thesis offers new and reliable insight into the language used to discuss mental illness in the UK press. Specifically, the research provides more specific insight into this area because, unlike previous research, it not only

[85] To my knowledge.

identifies features of and themes in news reports on mental illness (e.g. criminality) but also describes how these features and themes are linguistically manifested, e.g. through the use of certain naming practices. In addition to providing more nuanced information about news reports on mental illness, the research I report here has also provided information that may be useful for future anti-stigma initiatives that take account of linguistic evidence. For example, the research I reported in Chapter 8 on the common collocates of words that *Time to Change* have identified as contributing to stigma around mental illness (e.g. 'suffer') may be useful because it provides a means of identifying textual links between words deemed stigmatising, rather than seeing individual words as being inherently stigmatising. As a result of bringing attention to these textual links between linguistic forms that have been deemed problematic, this research provides a wider picture of stigma creation through *language* (e.g. texts) rather than individual linguistic forms (e.g. words). In addition, the analysis I report in Chapter 6 of this thesis, which I have been able to conduct due to the design of the MI 1984-2014 corpus, constitutes the first exploration of the terms 'mental health' and 'mental illness', whose meaning, I argued, have previously been taken for granted in the prior research (because such prior research does not privilege language as an object of study). The findings reported in Chapter 6 not only provide an insight into what 'mental health' and 'mental illness' mean across the time period covered by the corpus, but also provide predictive insight into what the nature of future lexical change will be in mental health and illness discourse (e.g. that 'wellbeing' will become more widely adopted to refer to *mental wellness*).

In this chapter, I provide a final discussion of my research findings and the implications of the research I have conducted in this thesis. In Section 10.1, I address the research questions listed in the Chapter 1 'Introduction'. In Section 10.2, I discuss the implications of the research reported in this thesis. Section 10.3 describes some potential limitations of this research, and in Section

10.5 I make suggestions for future research conducted in the area of representations of mental illness in language data. In Section 10.5, I conclude this thesis.

10.1. Fulfillment of research questions

I order to address how I have answered the research questions I outlined in Chapter 1, it will be useful to restate them here:

1. How are the terms 'mental illness' and 'mental health' used in the MI 1984-2014 corpus?
2. What linguistic strategies are used to name, label and describe people with mental illness in the MI 1984-2014 corpus?
   2.1. To what extent is person-first language present in the MI 1984-2014 corpus?
   2.2. What themes are present in the corpus for referring to people with mental illness?
3. What processes are associated with mental illness in the MI 1984-2014 corpus?
   3.1. What terms do the press use to refer to having mental illness?
4. Is the depiction of mental illness realistic?
   4.1. Are the symptoms of each disorder type (e.g. depressive illnesses) accurately portrayed in the press?

Throughout this thesis I have argued that no previous linguistic research has explored the language used to report on mental illness in the UK press (see Atanasova et al., 2019 for a linguistic analysis of arts initiatives in local newspapers reporting on mental health). As a result, the pool of knowledge in linguistics on this topic is limited. For this reason, my research questions were

designed to answer some fairly basic linguistic questions about the nature of the language used in the MI 1984-2014 corpus; for example, what 'mental health' and 'mental illness' mean. This question relates to RQ1.

In order to answer RQ1, I explored the use of 'mental health' and 'mental illness' across the time period covered by the corpus. I found that the frequency of both 'mental illness' and 'mental health' increased over time. I argued that this finding indicated that mental health and illness is an increasingly important social issue. I also found that the increase in 'mental health' was significantly higher than that of 'mental illness'. Closer inspection of 'mental health' and 'mental illness' through concordance and collocation analysis revealed that there appeared to be a semantic shift ongoing during the period (1984-2014), whereby 'mental health' was being used to refer to states of mental illness. I argued that this semantic shift was the result of socially-motivated, euphemistic language change. I argued that this language change was in accordance with existing theories of language change and euphemism, which indicated that the term 'mental illness' was taboo (resulting in euphemistic forms like 'mental health problem'). I also argued that the way the press use 'mental health' in the MI 1984-2014 corpus (particularly after c.2008) differed from the definition of 'mental health' given by the World Health Organisation (WHO). As a result, my research provides some preliminary insight into a definition of 'mental health' (in a UK context) that is enriched by contemporary language data. In addition to exploring the use of 'mental health' and 'mental illness' in the corpus, I reported linguistic evidence that showed that 'wellbeing' was used in the data to refer to states of mental wellness. I argued that the fact that 'wellbeing' was not textually associated with words like 'stigma' meant that 'wellbeing' has the potential to be adopted by anti-stigma initiatives to refer to *mental wellness*, given that 'mental health' is textually associated with 'mental illness' and therefore with stigma. Moreover, I made the case that, unlike 'mental health' which has a conventionalised negative

form ('mental illness'), 'wellbeing' does not have a conventionalised antonym and does not have a marked form to denote negation. This means that 'wellbeing' would not be susceptible to the same process of semantic shift that 'mental health' is undergoing.

Furthermore, I argued that the fact that 'mental health' and 'mental illness' are used in distinct ways has wider implications for the methodological best practice of studies whose aims are to model the semantic domain of MENTAL HEALTH AND ILLNESS. To summarise these findings and to answer RQ1, then, the first finding is that the corpus showed a developing tendency for the press to use the term 'mental health' over 'mental illness'. The second finding related to RQ1 is that 'mental illness' is a taboo term and that in order to discuss mental illness without using this term, 'mental health' is used to refer to states of mental illness. The third finding relating to RQ1 is that the way 'mental health' is used in MI 1984-2014 is distinct from definitions of 'mental health' given by some major mental health stakeholders, e.g. the WHO.

RQ2 was concerned with the naming practices used in the data to refer to people with mental illness. The reason for exploring the naming practices in the news reports was that naming practices have been identified "a powerful ideological tool [and] […] an accurate pointer to the ideology of the namer [because] different names for an object represent different ways of perceiving it" (Clark, 1992: 209). The two sub-questions of RQ2 related to whether person-first language was used in the corpus (RQ2.1) and what themes were present in the naming practices RQ2.2). In relation to RQ2, I identified that the salient head nouns for referring to people with mental illness are 'patient', 'victim' and 'sufferer'. I reported that the frequency of these terms over time was starting to plateau, which I argued gave some evidence for the fact that these terms are now established in the discourse of mental health and illness. I showed how 'patient', 'victim' and 'sufferer' carry different semantic associations that may contribute to stigma around people with mental illness. Specifically, I argued

that 'victim' is used to refer to people in irreversible contexts (e.g. a victim of suicide). As a result of this, I argued that 'victim' suggests that the illness the person is a 'victim' of is not treatable, which may contribute to stigma around mental illness. I also showed linguistic evidence from the corpus to argue that 'patient' is pathologised, which marks this term as being abnormal in some way. This is despite the fact that 'patient' is not identified as a problematic term by *Time to Change*. Moreover, I showed through an analysis of 'sufferer', 'patient' and 'victim' in the illness subcorpora how the use of these three labels is patterned in the MI 1984-2014 corpus. I found that 'patient' (which implies pathology) was more likely to occur in reference to psychosis and schizophrenia, which I argued gave linguistic evidence for the notion that psychotic disorders like schizophrenia are more stigmatised than other illnesses (e.g. Mann & Himlein, 2004) because they are pathologised in a way that other mental illnesses are not. In addition to this finding, I also reported that 'victim' is most often used in the PTSD corpus. In contrast to 'patient' and 'victim', the label 'sufferer' is not markedly more frequent in one subcorpus. The label 'sufferer' is most frequent in the SAD, OCD and SocialPhobia subcorpora.

In relation to RQ2.1, I found that person-first language is present in the MI 1984-2014 corpus; however, it is infrequent. I showed that overall, the identity-first variant of naming (e.g. a schizophrenic) is much more common than person-first forms. I also found that despite the low frequency of person-first terms, there is an overall positive trend for person-first forms which indicates that the press are increasingly adhering to prescribed person-first forms. I also found in my analysis of person-first forms that person-first language occurs in contexts that are supportive of people with mental illness. However, I also found that naming practices that have been identified as problematic by *Time to Change* occur additionally in contexts that are supportive of person-first forms. I argued that this demonstrates that no

linguistic form is inherently stigmatising and that the relationship between person-first language and positive representations of mental illness (e.g. supportive contexts) is one of correlation and not causation. This finding also provides more evidence for the benefits of linguists working with anti-stigma advocates (because linguists would not claim that a linguistic form is inherently stigmatising and linguists would be able to provide linguistic evidence for such claims).

In my analysis of person-first forms, I also reported the tendency in the corpus to refer to people with mental illness in the plural form, e.g. 'people with depression' over 'person with depression'. This leads me on to the findings I made in response to RQ2.2 ("What themes are present in the corpus for referring to people with mental illness?"). In my analysis of the salient naming practices in the corpus, I found that a key theme was labelling people with mental illness as quantities. I presented linguistic evidence from qualitative and quantitative analysis of the corpus to show that the press routinely present people as statistics. I argued that the overrepresentation of statistics in news reports on mental illness is a device used by journalists to sensationalise news stories pertaining to mental illness. Moreover, I argued that the overrepresentation of statistics and numbers as well as verbs denoting rising numbers contributes to what Fowler (1991) terms 'press hysteria' via the "rhetoric of quantification" (Fowler, 1991) in news reports on mental illness. Furthermore, using evidence from the corpus, I demonstrated how news reports contribute to the portrayal of mental illness as a "moral panic' (Cohen, 1973). Furthering Fowler's (1991) notion of a 'rhetoric of quantification', I argued that naming people with mental illness as statistics contributes to a 'rhetoric of depersonalisation" whereby the systematic representation of people as numbers backgrounds the experiences of individuals. I argued that the effect of this depersonlisation is that readers will be less likely to empathise with individual people and their experiences of mental illness.

RQ3 was concerned with the salient transitivity processes in the corpus. The reason for dedicating a research question to transitivity processes was that the ways people are described as having mental illness has also garnered attention from mental health advocates promoting prescribed linguistics forms (e.g. whether a person 'suffers' from a mental illness, or 'experiences' a mental illness). I designed RQ2.1 ("What terms do the press use to refer to having mental illness?") to explore how the press discuss having mental illness in the MI 1984-2014 corpus.

In my analysis of the data in response to RQ3, I found that 'suffer' is present in a third of the prototypical text sample. Due to the fact that 'suffer' has been identified by anti-stigma initiatives such as *Time to Change* as a problematic way to describe having mental illness, I analysed the lexicogrammatical function of 'suffer' in the corpus. I also analysed the lexicogrammatical function of the verb 'experience' (e.g. 'she experienced depression') due to the fact that *Time to Change* identify 'experience' as a preferred form. In my analysis of the two terms, I found that 'suffer' is negatively-valanced, and occurs in contexts that remove agency from the 'sufferer'. As a result of this, 'suffer' implies that people are not in control of their illness, which I argued is an inaccurate depiction of many people with mental illness. In contrast, 'experience' was neither negatively- or positively-valanced, but rather conveyed a sense of 'learned' (e.g. "learned from experience"). Moreover, using evidence from the BNC, I showed that 'experience' encodes agency; i.e. an experience has to be experienced by an 'experiencer'. I also showed that 'suffer' is used to refer to unbounded (e.g. not temporally-fixed) phenomena, whereas 'experience' is temporally-fixed. Through an exploration of the collocates of 'suffer', I also showed that, unlike 'experience', 'suffer' collocates with other linguistic forms that are deemed problematic by *Time to Change*, such as 'victim'. Taken together, I argued that my findings provided linguistic evidence for the claim that 'experience' is a

better term to describe having mental illness than 'suffer' because 'suffer' is routinely used in negative contexts, which suggests that the 'sufferer' is not in control of their illness and is textually associated (through collocation) with other problematic forms. Based on this finding, I suggested that an increase in the use of 'experience' to describe having mental illness would not just result in a reduction in the use of 'suffer', but could also result in a reduction of other problematic forms that collocate with this verb. This finding provides another example of how collaboration between mental health advocates and linguists could pave new, evidence-based linguistic guidelines (e.g. exploring collocates of problematic forms).

As part of my exploration of the data in response to RQ3, I also explored the use of 'suffer' and 'experience' in self-reflexive contexts, e.g. "I suffer", "I experience". The reason for doing this was to further explore the linguistic basis for prescribing 'experience' and identifying 'suffer' as problematic, because if 'suffer' is problematic for people with mental illness, then one would not expect frequent self-reflexive use of 'suffer'. I found that despite initial results suggesting that people with mental illness do use 'suffer' reflexively, it is actually the case that '[I + 'experience'] is proportionally four times more frequent in the corpus that '[I + suffer]'. The reason for this is that '[I + suffer]' often occurs in reported speech, e.g. "the doctor decided I was suffering with depression".

In my exploration of 'experience' and 'suffer' I also found there to be conceptual differences in how 'suffer' and 'experience' are used in the corpus and in the BNC which I argued was related to diagnostic status. In a concordance analysis of the data, I found that 'experience' is typically used to describe symptoms of an illness (which may not have been diagnosed), whereas 'suffer' is the process used to describe having a diagnosed illness, e.g. "she experienced low mood and fatigue" vs. "she suffered with depression". I provided further evidence for this interpretation by showing that 'suffer'

occurs with subject collocates featuring identify-first labels such as 'schizophrenic' whereas 'experience' never occurs in the corpus with a subject collocate that features an identity-first form (meaning that labels in subject position that encode diagnosis do not collocate with 'experience').

I also found that unlike 'experience', 'suffer' is used (in the MI 1984-2014 corpus and in the BNC) in both mental processes, e.g. "she suffered from depression" (taking a Senser, Process, Phenomenon configuration), and material processes with inanimate actor, e.g. "the fence suffered damage from the wind" (taking a Goal, Circumstance, Actor configuration). As a result of this finding, I argued that the negation of agency of 'suffer' I discussed previously may be exacerbated by its association with non-conscious entities (e.g. 'the fence' in the example above). I also argued that the data demonstrates a further conceptual difference: a difference between transitive and intransitive 'suffer'. Using evidence from the MI 1984-2014 corpus and the BNC, I argued that there is a tendency in the data for intransitive uses of 'suffer', e.g. 'suffer with' to occur with mental states (e.g. she suffered with depression), whereas transitive uses of 'suffer', occur most often with physical states (e.g. she suffered a broken leg). I argued that this finding gives some linguistic evidence for a perceived duality (conveyed in language) between mental and physical illness. I argued that the lexicogrammatical analysis conducted in response to RQ3 enriches existing dictionary definitions of 'suffer' in addition to existing semantic explications of 'suffer' (e.g. Shweder 2003; Wierzbicka, 2016). I reported that in answer to RQ3 ('What processes are associated with mental illness?'), 'suffer' is a salient process in news reports on mental illness whereas 'experience' is less common. I also argued that, for the reasons I identified in Chapter 8 and report here, 'experience' is a more appropriate term for the press to use than 'suffer'.

The last research questions I addressed I this thesis was RQ4 ("Is the depiction of mental illness realistic") and RQ4.1 ("Are the symptoms of each

disorder type (e.g. depressive illnesses) accurately portrayed in the press?"). The reason for asking this question of the data was that previous research has largely neglected to explore whether the symptoms of mental illnesses are accurately portrayed in news reports on mental illness. Moreover, those studies that do explore the whether journalists reference symptoms in news reports only analysed whether any symptoms were present in the articles and not how the symptoms were described, or whether the descriptions were accurate or not. (Wahl et al. 2002).

In order to answer RQ4 and RQ4.1, I used keyness analysis to investigate the statistically overused keywords and key semantic domains in each of the subcorpora that relate to a disorder type, e.g. depressive disorders. I also enriched this corpus linguistic analysis with the analysis of the most prototypical article of each illness subcorpora (the most prototypical text is the text in the subcorpus that contains the most keywords in comparison with a reference corpus). I found that overall the representation of symptoms in the subcorpora is mixed. In order to answer RQ4.1 in detail, I report the specific findings for each disorder type below:

- **Trauma disorders**

  I found that overall, many of the symptoms of PTSD were present in the PTSD corpus; however, the contexts in which these symptoms are commonly reported were too specific to accurately represent PTSD. Specifically I found the symptoms of PTSD in combat situations were overrepresented whereas the symptoms of PTSD in reference to birth trauma were underrepresented. I argued that the tendency in news reports on PTSD to focus on combat situations only constitutes a disparity between the reality of the condition and the representation of it in the press.

- **Dissociative disorders**

  I found that compared with other disorder types, articles on dissociative disorders are much less common in the corpus, which limited the possible research findings. I found that articles that report on DID, although few in number, do generally include a discussion of symptoms, although the symptoms reported are usually those experienced by a single person. As a result of this, a general description of the symptoms of DID are not present in the corpus.

- **Bipolar disorder**

  I found that despite the fact that some of the commonly known symptoms of bipolar disorder (such as high and low mood) are represented, the symptoms of bipolar disorder are inaccurate overall. The reason for this is that news reports neglect to represent the psychotic symptoms of bipolar disorder such as voice hearing and hallucinations. Moreover, there is a tendency in the BipolarDisorder corpus to report on celebrities' experiences of bipolar disorder and not the symptoms of bipolar disorder generally.

- **Obsessive compulsive disorders**

  I found that news reports on OCD did not include the full range of OCD symptoms. Furthermore, I found that news reports on OCD overrepresent the compulsion side of OCD (e.g. washing and cleaning hands) and therefore underrepresent the obsessive side of OCD (e.g. sexual intrusive thoughts). I argued that the overrepresentation of the obsession side of OCD reinforces common misconceptions that OCD is only about cleaning.

- **Psychotic disorders**

  I found that a common feature of news reports on psychotic disorders is reference to drugs (e.g. 'skunk' and 'cannabis'). I argued that the overrepresentation of words related to drugs is indicative of a tendency in the press to refer to the causes of psychotic disorders and not the symptoms of them. I showed using evidence from the MI 1984-2014 corpus that lexis relating to common symptoms of psychotic disorders (e.g. 'delusions' and 'hallucinations') are comparatively low. I showed that the occurrence of the lexical item "delusion*" is less frequent in the corpus (per million words) than lexical items associated with isolated terrorist incidents, such as 'Breivik'. An analysis of the most prototypical text in the psychosis corpus revealed that whilst symptoms of the psychotic disorder schizophrenia are included in the article, they are reported in reference to a violent attack. Due to the fact that the articles often do not generally feature symptoms, and when they so they are used in reference to violent crime, I argued that the symptoms of psychotic disorders are not accurately portrayed in news reports.

- **Anxiety disorders**

  I found that in news reports on anxiety disorders (e.g. specific and complex phobias), complex phobias (such as social phobia and agoraphobia) are underrepresented. I showed that keyness analysis reveals a fear of flying to be a key theme in the anxiety corpus which relates to a specific phobia. I argued that the lack of representation of symptoms of anxiety disorders could be due to the fact that many of the symptoms of anxiety disorders are also symptoms of other mental illnesses and therefore would not be identified by Wmatrix as key.

- **Eating disorders**

  I found that the articles do not accurately represent the symptoms of eating disorders. I argued that this was due to the fact that news reports focus on weight loss and physical appearance (particularly in women) which does not reflect that eating disorders are a self-esteem issue. Moreover, I argued that key words pertaining to women's clothes sizes (e.g. size 10), showed that eating disorders are portrayed as a female illness and not one that affects men. I showed that the representation of eating disorders in the press is in contrast to the statistics on eating disorders which show that a quarter of all people with eating disorders are men.

- **Depressive disorders**

  I found that the news reports of depressive disorders did include some of the symptoms of depression, e.g. reduced libido and loss of appetite. I also found that discussion of eating disorders also occurs in the depression corpus (people with eating disorders can also experience depression). Furthermore, I found that, like some other illnesses in the corpus (e.g. Bipolar disorder), depression is commonly reported in the context of celebrity sufferers.

In addition to exploring the representation of symptoms using keyness analysis, I also showed how collocation is revealing of the symptoms discussed in the corpus which were not apparent in the keyness analysis. Moreover, I showed that collocation analysis can provide insight into overlap in the symptoms represented between illness types. I argued that the fact that some celebrity names are statistically significant in some illness subcorpora (e.g. depression and bipolar) and not others (psychotic disorders) suggests that psychotic disorders are not illnesses that people align with. I argued that this

finding is a linguistic indication (i.e. through keyness analysis) that psychotic disorders are more stigmatised than other illnesses that do have celebrity spokespeople.

In the next section, I outline the implications of the research reported in this thesis.

10.2. Implications of this research

As I previously stated, the research I have reported in this thesis constitutes the first, large-scale linguistic analysis of UK news reports on mental health generally (i.e. not motivated by a particular theme or focused on a particular illness). As a result of this, the first implication of this work is that it creates a foundation level of knowledge for this area. In addition to this, the MI 1984-2014 corpus is the first corpus containing news reports on mental illness that is optimised for the analysis of synchronic and diachronic variables (and depending on the research question, provides relevant and representative reference corpora). As a result, the MI 1984-2014 corpus constitutes a significant resource for the field. The first implication of my research, then, is that it contributes new knowledge and new resources to the field.

The second implication of this research is that the research findings I have reported offer a much more specific and nuanced understanding of how mental illness is reported in the UK press, from the very labels used to describe mental health and mental illness (Chapter 6) to how symptoms are reported. My research has built on the significant body of research in psychiatry and has demonstrated that the analysis of language as an object of study in its own right can provide never-before-reported findings into various areas of mental illness research such as stigma, the depiction of mental illness generally, and the depiction of specific mental illnesses. Due to the fact that several parts of this

research have been concerned with the semantic properties of words within the semantic field of mental health and illness (and providing a usage-based, nuanced definition of words within that semantic field), my research also has potential lexicographical implications (e.g. the semantic shift occurring in the term 'mental health' reported in Chapter 6, and the more nuanced meaning of transitive and intransitive uses of 'suffer' reported in Chapter 8). As a result, the research reported in this thesis has implications in linguistics, but also in psychiatry.

Furthermore, my research has methodological implications for the field of corpus linguistics, specifically in the area of corpus construction. I argue that the corpus construction procedure I outline in Chapter 5 is innovative as it takes into account the interpretative status of search terms. Specifically, I argue that the search terms used to construct corpora for the analysis of ideology are not interpretatively neutral. I outlined a series of questions in Chapter 5 (Section 5.1) that the researcher should be able to answer when constructing corpora for this purpose. Moreover, I argued in Chapter 3 ('Analytical Methods Part 1: Corpus Linguistics') that an innovative use of n-gram analysis is to use n-grams to identify potential data skew in determining the representativeness of a corpus (rather than as an analytical tool). My research , then, has clear methodological implications.

Outside of linguistics, my research has implications within the medical humanities more generally because I have shown that the terms 'mental health' and 'mental illness' do differ in their linguistic contexts. Therefore research conducted in psychiatry that uses these labels interchangeably to collect press data is problematic. That is, the insights from my thesis may be used to improve working methods in the medical humanities.

Beyond these academic implications, there are also practical implications for campaigns such as Time to Change, and for journalists writing about the issues discussed in this thesis. My research is well placed to have implications

outside academia (e.g. societal impact) because of its relevance to public institutions such as anti-stigma organisations and dictionaries. For instance, Time to Change is a campaign that is particularly focused on tracking change over time. However, the focus in mental health research generally is not on language, which limits the level of nuanced understanding of change that such research can provide. Because my own research does prioritise language as an object of study in itself, and utilizes computational methods, I am able to provide an insight into changes in language use that go beyond simple value judgements of whether a word is inherently positive or negative. For example, my analysis of 'wellbeing' and 'mental health' in Chapter 6 demonstrates that change is incremental and inevitable and that language is much more flexible than advocates of prescribed forms suggest. These findings could be integrated into language awareness exercises for writers on the topic of mental health and could be used to assist campaigners in developing better working practices and more effective campaign strategies.

In the next section, I describe some limitations of this research.

10.3. Limitations

The fist limitation of this research is that the data is monomodal. Nexis does not save the images that accompany the text in an article, nor does it maintain the formatting of articles, e.g. font size differences between the main body of an article and the headline. In addition, the mainstream corpus tools (e.g. Sketch Engine, Wmatrix, Antconc) do not yet have the functionality to analyse the visual aspects of texts. As a result, even if Nexis did retain the visual features of the article, I would not have been able to analyse those aspects using the methods adopted in this thesis. The monomodal nature of my analysis constitutes a limitation because previous research has shown that the

multimodal aspects of texts are meaningful and contribute to ideology (e.g. Harvey, 2013; Harvey & Brookes, 2017; Lirola & Chovanec, 2012).

A further limitation of this research is that it is discourse-specific, i.e. it is limited to the representation of mental illness in *newspaper data*. As a result of this, the research cannot account for how people speak or write about mental health outside of newspaper discourse, which is heavily edited and intended for a wide public audience. Due to this, newspaper discourse cannot account for how individuals in UK society discuss mental illness, which (if the research is interested in stigma, for example) could possibly be more stigmatising due to the fact that it is unedited and private. With these points in mind, I argue that newspaper discourse is still a useful data type for investigating commonly-held beliefs about mental illness, and what is deemed appropriate or problematic in relation to mental illness at a given point in time.

In addition to the limitations listed above, a further limitation of research that combines the analysis of big data (like the MI 1984-2014 corpus) with in-depth qualitative analysis is the issue of striking a balance between the size of the data and the analytical depth into which a researcher can go. In this thesis I have utilised the size and scope of the MI 1984-2014 corpus to show, for example, semantic change over time. I have also used the corpus to explore some very small-scale qualitative questions; for example, the semantic difference between the lexical items 'suffer' and 'experience'. As a result of giving equal attention to small- and large-scale linguistic phenomena, my research here is perhaps open to the criticism that it does not explore the corpus in its entirety.

Notwithstanding whether exploring a corpus in its linguistic entirety is even possible, I argue that the balance between the size of the data and analytical depth is a strength as well as a limitation of this work because it allows me to focus on the data at varying levels of magnification. Throughout this thesis, I have used this affordance of corpus methods as a means of research

triangulation. One example of how I have done this is by demonstrating that the labels 'mental health' and 'mental illness' are changing over time by showing this change through several analytical vantage points, e.g. through collocation analysis, concordance analysis and frequency analysis. Whilst in isolation these individual analyses are not detailed enough to support this claim, together they provide strong evidence for this semantic shift. In this way, I have endeavored to ensure that claims made on the basis of qualitative analyses of small sections of the corpus are supported by quantitative evidence from the corpus as a whole, and vice versa.

In the next section, I make some suggestions for future research.

10.4. Suggestions for future research

A large portion of this thesis has been dedicated to discussion of stigma and stigmatising linguistic forms. In my analysis, I used the MI 1984-2014 corpus to provide an initial exploration of the linguistic basis for some of the linguistic forms that have been prescribed by anti-stigma initiatives like *Time to Change* (e.g. using 'experience' over 'suffer', avoiding person-first language, and avoiding calling people with mental illness 'sufferers'). The reason for doing this was that the prescribed forms for discussing mental illness in the press do not seem to be informed by any linguistic analysis. A future research avenue in this area would be to investigate whether there is any linguistic basis for these prescriptions from a pragmatic perspective, informed by the opinions of people with mental illness - for example, through interviews that ask informants to explore whether they find a particular linguistic form stigmatising in real-world language data (e.g. newspaper data). To my knowledge only one study explores terms that are deemed stigmatising using people with mental illness as participants, and that study was specific to schizophrenia and was

conducted by a psychiatrist (Haghighat, 2008). In his research, Haghighat (2008) actually found that people with schizophrenia preferred the label 'schizophrenic' over person-first forms, although the participants were asked to report on words only and not on those words used in any naturalistic context. Knowing this, there is substantial scope for an exploration of prescribed linguistic forms using participants with mental illness and using real-world language data.

The future research avenue I just suggested relates to a weakness in corpus linguistic analysis more generally, which is that it is focused on the production of texts rather than the comprehension of them (i.e. the analyst does not know what effect the texts *actually* have on readers). A future research avenue related to addressing this weakness would be to conduct psycholinguistic experiments to explore the cognitive basis for some problematic forms - for example, using eye-tracking methods (e.g. Conklin & Schmitt, 2008) to investigate the psychological validity of the collocations I identified in Chapter 8 (e.g. that 'suffer' collocates with 'victim' and both are deemed problematic by *Time to Change*). Furthermore, new methodologies for investigating the psychological validity of collocation are being developed in corpus linguistics and cognitive neuroscience, e.g. using electroencephalogram (EEG) tests (Hughes, 2018).

10.5. Conclusion

In this chapter, I restated the motivation for this research and showed how it addresses gaps in the existing research into the discursive construction of mental illness. I have revisited in turn the research questions that guided the analysis reported in this thesis and reported my research findings pertaining to these questions. I have argued that the research reported in this thesis has implications in various ways. First it contributes a more nuanced and evidence-

based understanding of the language used to discuss mental illness. Second, it has methodological implications for corpus linguistics. Third, it has implications outside of academia in that it contributes never-before-reported findings that have the potential to have clear societal impact. The language we use to write about mental health and mental illness is of the utmost importance. My hope is that this thesis will provide the basis for further exploration of this important topic.

References

Adolphs, S., Brown, B., Carter, R., Crawford, P. and Sahota, O. (2004) 'Applying corpus linguistics in a health care context', *Journal of Applied Linguistics* 1(1): 9-28.

Allan, K. and Burridge, K. (2006) *Forbidden Words: Taboo and the Censoring of Language*. Cambridge: Cambridge University Press.

Angermeyer, M. C., Dietrich, S., Pott, D. and Matschinger, H. (2005) 'Media consumption and desire for social distance towards people with schizophrenia', *European Psychiatry* 20(3): 246-250.

Angermeyer, M.C. and Matschinger, H. (2005) 'Labeling—stereotype—discrimination', *Social Psychiatry and Psychiatric Epidemiology* 40(5): 391-395.

Anorexia and Bulimia Care (2019) *Statistics* [online] Available at: <http://www.anorexiabulimiacare.org.uk/about/statistics> [Accessed 2 Aug 2019].

Antaki, C. (2007). Mental-health practitioners' use of idiomatic expressions in summarising clients' accounts', *Journal of Pragmatics* 39(3): 527-541.

Anthony, L. (2017) *AntConc* 3.4.4. Tokyo, Japan: Waseda University. Available at: http://www.laurenceanthony.net. (Accessed 1 September 2018)

Anthony, L. and Baker, P. (2015) 'ProtAnt: A tool for analysing the prototypicality of texts', *International Journal of Corpus Linguistics* 20(3): 273-292.

Aoki, A., Aoki, Y., Goulden, R., Kasai, K., Thornicroft, G. and Henderson, C. (2016), 'Change in newspaper coverage of schizophrenia in Japan over 20-year period', *Schizophrenia Research* 175(1-3): 193-197.

APA (2019) What is mental illness? [online]. Available at: https://www.psychiatry.org/patients-families/what-is-mental-illness [Accessed 21 March 2019].

Archer, D. (ed.) (2009) *What's in a Word-List?: Investigating Word Frequency and Keyword Extraction*. Ashgate Publishing, Ltd.

Ariel, M. (2008) *Pragmatics and Grammar.* Cambridge: Cambridge University Press.

Atanasova, D., Koteyko, N., Brown, B. and Crawford, P. (2019) 'Representations of mental health and arts participation in the national and local British press, 2007–2015', *Health* 23(1): 3-20.

Athanasopoulou, C. and Välimäki, M. (2014) '"Schizophrenia" as a metaphor in Greek newspaper websites', *ICIMTH* 202: 275-278.

Atlas, J. D. and Levinson, S. C. (1981) 'It-clefts, informativeness and logical form: radical pragmatics', in Cole, J. (ed.) *Radical Pragmatics*, pp. 1-62. New York: Academic Press.

Bailey, R. W. (1969) 'Statistics and style: a historical survey', in Dolezel, L. and Bailey, R. W. (eds) *Statistics and Style*, pp. 217-236. New York: Elsevier.

Baker, C. (2018) 'Mental health statistics for England: prevalence, services and funding', *Second Reading. London: House of Commons Library Briefing*.

Baker, P. (2016) 'The shapes of collocation', *International Journal of Corpus Linguistics* 21(2): 139-164.

Baker, P. and Levon, E. (2015) 'Picking the right cherries? A comparison of corpus-based and qualitative analyses of news articles about masculinity', *Discourse and Communication* 9(2): 221-236.

Baker, P., Gabrielatos, C. and McEnery, T. (2012) 'Sketching muslims: a corpus driven analysis of representations around the word 'muslim' in the British press 1998–2009', *Applied Linguistics* 34(3): 255-278.

Baker, P., Gabrielatos, C., Khosravinik, M., Krzyżanowski, M., McEnery, T. and Wodak, R. (2008) 'A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press', *Discourse and Society* 19(3): 273-306.

Barnbrook, G. (1996) *Language and Computers*. Edinburgh: Edinburgh University Press.

Barnish, M. (2014) 'A quantitative content analysis of person-first language use in healthcare research, healthcare practice and by support groups for people with disabilities ', *Open Journal of Modern Linguistics* 4: 505-511.

Bateman, J. A. (2017) 'The place of systemic functional linguistics as a linguistic theory in the twenty-first century', in Bartlett, T. and O'Grady, G. (eds) *The Routledge Handbook of Systemic Functional Linguistics*, pp. 35-50. Abingdon: Routledge.

Battistella, E. (1990) *Markedness: The Evaluative Superstructure of Language*. New York: SUNY Press.

Bednarek, M. (2006) *Evaluation in Media Discourse: Analysis of a Newspaper Corpus*. London: Continuum.

Bednarek, M. (2008) 'Semantic preference and semantic prosody re-examined', *Corpus Linguistics and Linguistic Theory* 4(2): 119-139.

Bednarek, M. and Caple, H. (2014) 'Why do news values matter? Towards a new methodological framework for analysing news discourse in Critical Discourse Analysis and beyond', *Discourse and Society* 25(2): 135-158.

Bednarek, M. and Caple, H. (2017) *The Discourse of News Values: How News Organizations Create Newsworthiness*. Oxford: Oxford University Press.

Bell, A. (1991) *The Language of News Media* .Oxford: Blackwell.

Bell, A. (1994) 'Telling stories', in Graddol, D. and Boyd-Barrett, O. (eds) *Media Texts: Authors and Readers*, pp.100-118. London: Multilingual Matters.

Berger P. L. and Luckmann, T. (1966) *The Social Construction of Reality: A Treatise in The Sociology of Knowledge*. New York: First Anchor.

Bilić, B. and Georgaca, E. (2007) 'Representations of "mental illness" in Serbian newspapers: a critical discourse analysis', *Qualitative Research in Psychology* 4(1-2): 167-186.

Bipolar UK (2019 *Bipolar - the facts* [online]. Available at: <https://www.bipolaruk.org/faqs/bipolar-the-facts> [Accessed 2 Aug 2019].

Björkman, T., Svensson, B. and Lundberg, B. (2007) 'Experiences of stigma among people with severe mental illness. Reliability, acceptability and construct validity of the Swedish versions of two stigma scales measuring devaluation/discrimination and rejection experiences', *Nordic Journal of Psychiatry* 61(5): 332-338.

Blank, A. (1999) 'Why do new meanings occur? A cognitive typology of the motivations for lexical semantic change', in Blank, A. and Koch, P (eds) *Historical Semantics and Cognition*, pp.61-83. Berlin: Mouton de Gruyter.

Bloor, T. and Bloor, M. (2013) *The Functional Analysis of English: A Hallidayan Approach*. Abingdon: Routledge.

Blumer, H. (1969) *Symbolic Interactionism: Perspective and Method*. Berkeley: University of California Press.

Boas, F. (1917) 'Introduction', *International Journal of American Linguistics* 1(1): 1-8. [Reprinted in Boas, D. (1940) *Race, Language and Culture*, pp. 199-210. The Free Press; New York.]

Bondi, M. and Scott, M. (eds) (2010) *Keyness in Texts*: Amsterdam: John Benjamins.

Bowen, M. L. (2016) 'Stigma: content analysis of the representation of people with personality disorder in the UK popular press, 2001–2012', *International Journal of Mental Health Nursing* 25(6): 598-605.

Brinton, L. J. and Traugott, E. C. (2005) *Lexicalization and Language Change*. Cambridge: Cambridge University Press.

Britain, D. (2018) 'Paris: a sociolinguistic comparative perspective', *Journal of French Language Studies* 28(2): 291-300.

Brookes, G. and Harvey, K. (2017) 'Just plain Wronga? A multimodal critical analysis of online payday loan discourse', *Critical Discourse Studies* 14(2): 167-187.

Bucholtz, M. and Hall, K. (2005) 'Identity and interaction: A sociocultural linguistic approach', *Discourse Studies* 7(4-5): 585-614.

Buerki, A. (2016) 'Formulaic sequences: a drop in the ocean of constructions or something more significant?', *European Journal of English Studies* 20(1): 15-34.

Burr, V. (1995) *Social Constructionism*. London: Routledge

Burridge, K. (2012) 'Euphemism and language change: the sixth and seventh ages', *Lexis: Journal in English Lexicology* (7): 65:92.

Butler, C. (1985) *Systemic Linguistics: Theory and Applications*. London: Batsford Academic and Educational.

Butler, C. S. (2003) '*Structure and Function: A Guide to Three Major Structural-Functional Theories: Part 2: From Clause to Discourse and Beyond*. Amsterdam: John Benjamins.

Cameron, D. (2012) *Verbal Hygiene*. Abingdon: Routledge.

Canning, P. (2013) 'Functionalist Stylistics' in Burke, M. (ed.) *The Routledge Handbook of Stylistics,* pp.45-67. Abingdon: Routledge.

Caple, H. (2013) 'Competing for coverage: Exploring emerging discourses on female athletes in the Australian print media', *English Text Construction* 6(2): 271-94.

Carston, R. (1998) 'Informativeness, relevance and scalar implicature' in Carston, R. and Uchida S. (eds) *Relevance Theory: Applications and implications*, pp. 179-238. Amsterdam: John Benjamins.

Chan, S. K. W., Tam, W. W. Y., Lee, K. W., Hui, C. L. M., Chang, W. C., Lee, E. H. M. and Chen, E. Y. H. (2016) 'A population study of public stigma about psychosis and its contributing factors among Chinese population in Hong Kong', *International Journal of Social Psychiatry* 62(3): 205-213.

Chapman, S. (2011) *Pragmatics*. London: Palgrave Macmillan

Chopra, A. K. and Doody, G. A. (2007) 'Schizophrenia, an illness and a metaphor: analysis of the use of the term "schizophrenia" in the UK national newspapers', *Journal of the Royal Society of Medicine* 100(9): 423-426.

Clark, B. (2016) 'Relevance theory and language change', *Lingua* 175: 139-153.

Clark, H. H. and Lucy, P. (1975) 'Understanding what is meant from what is said: A study in conversationally conveyed requests', *Journal of Verbal Learning and Verbal Behaviour* 14(1): 56-72.

Clark, K. (1992) 'The linguistics of blame: Representations of women in the Sun's reporting of crimes of sexual violence' in Toolan, M. (ed.) *Language, Text and Context: Essays in Stylistics*, pp.208-224. London: Routledge.

Clarke, A. (2005) *Grounded Theory after the Postmodern Turn*. California: Sage.

Clement, S. and Foster, N. (2008) 'Newspaper reporting on schizophrenia: a content analysis of five national newspapers at two time points', *Schizophrenia Research* 98(1-3):178-183.

Cohen, L. (1995) 'Toward an anthropology of senility: anger, weakness, and Alzheimer's in Banaras, India', *Medical Anthropology Quarterly* 9(3): 314-334.

Cohen, S. (1973) *Folk Devils and Moral Panics*. St. Albans: Paladin.

Conboy, M. (2007) *The Language of the News*. Abingdon: Routledge.

Conklin, K. and Schmitt, N. (2008) 'Formulaic sequences: are they processed more quickly than nonformulaic language by native and nonnative speakers?', *Applied Linguistics*, 29(1): 72-89.

Conrad, P. and Barker, K. K. (2010) 'The social construction of illness: Key insights and policy implications', *Journal of Health and Social Behavior* 51(1): 67-79.

Corker, E., Hamilton, S., Robinson, E., Cotney, J., Pinfold, V., Rose, D. and Henderson, C. (2016) 'Viewpoint survey of mental health service users'

experiences of discrimination in England 2008–2014', *Acta Psychiatrica Scandinavica* 134: 6-13.

Corrigan, P. W. (1998) 'The impact of stigma on severe mental illness', *Cognitive and Behavioral Practice* 5(2): 201-222.

Corrigan, P. W. (2002) 'Empowerment and serious mental illness: treatment partnerships and community opportunities', *Psychiatric Quarterly* 73(3): 217-228.

Corrigan, P. W. and Watson, A. C. (2002) 'The paradox of self-stigma and mental illness', *Clinical Psychology: Science and Practice* 9(1): 35-53.

Corrigan, P. W., Watson, A. C., Gracia, G., Slopen, N., Rasinski, K. and Hall, L. L. (2005) Newspaper stories as measures of structural stigma. *Psychiatric Services 56*(5), 551-556.

Coulmas, F. (2016) 'Sociolinguistics and the English writing system', in Cook, V and Ryan, D. (eds) *The Routledge Handbook of the English Writing System*, pp.281-294. Abingdon, Oxon: Routledge

Coulthard, M. and Johnson, A. (2010) *The Routledge Handbook of Forensic Linguistics*. Abingdon: Routledge.

Coverdale, J., Nairn, R. and Claasen, D. (2002) 'Depictions of mental illness in print media: A prospective national sample', *Australian and New Zealand Journal of Psychiatry* 36(5): 697-700.

Culpeper, J. (2009) 'Keyness: words, parts-of-speech and semantic categories in the character-talk of Shakespeare's Romeo and Juliet', *International Journal of Corpus Linguistics* 14(1): 29-59.

Danielsson, P. (2013) 'What constitutes a unit of analysis in language?', *Linguistik Online* 31(2): 17-24.

Deamer, F. and Hayward, M. (2018), 'Relating to the speaker behind the voice: what is changing?', *Frontiers in Psychology* 9(11): 1-8.

Demjén, Z., Marszalek, A., Semino, E. and Varese, F. (2019), 'Metaphor framing and distress in lived-experience accounts of voice-hearing' *Psychosis*, 11(1): 16-27.

Dietrich, S., Heider, D., Matschinger, H.and Angermeyer, M. C. (2006) 'Influence of newspaper reporting on adolescents' attitudes toward people with mental illness', *Social Psychiatry and Psychiatric Epidemiology* 41(4): 318-322.

Dillon, G., Doyle, A., Eastman, C., Silberstein, S. and Toolan, M. (1993) 'Language and power in critical linguistics', *International Journal of the Sociology of Language* 103: 185-96.

Duckworth, K., Halpern, J. H., Schutt, R. K. and Gillespie, C. (2003) 'Use of schizophrenia as a metaphor in US newspapers', *Psychiatric Services* 54(10): 1402-1404.

Dunmire, P. L. (2005) 'Preempting the future: Rhetoric and ideology of the future in political discourse', *Discourse and Society* 16(4): 481-513.

Durrant, P.and Doherty, A. (2010) 'Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming', *Corpus Linguistics and Linguistic Theory* 6(2): 125-155.

Egbert, J. and Schnur, E. (2018) 'The role of the text in corpus and discourse analysis' in Taylor, C. and Marchi, A (eds) *Corpus Approaches to Discourse. A Critical Review*, pp.158-170. Abingdon: Routledge.

Eisenberg, L. (1977) 'Disease and illness distinctions between professional and popular ideas of sickness', *Culture, Medicine and Psychiatry* 1(1): 9-23.

Elizabeth, T.and Dasher, C. R. B. (2002) *Regularity in Semantic Change*. Cambridge: Cambridge University Press.

Enkvist, N. E. (1973) *Linguistic Stylistics*. The Hague: Mouton.

Fairclough, N. (1989) *Language and Power*. London: Longman.

Fairclough, N. (1992) *Discourse and Social Change*. Cambridge: Polity Press.

Fairclough, N. (1995) *Media Discourse*. London: Arnold.

Fairclough, N. (2003) *Analysing Discourse: Textual Analysis for Social Research*. Abingdon: Routledge

Fairclough, N. (2013) *Critical Discourse Analysis: The Critical Study of Language*. Abingdon: Routledge.

Fawcett, R. P. (1987) 'The semantics of clause and verb for relational processes in English' in Halliday, M. A. K. and Fawcett, R. P. (eds) *New Developments in Systemic Linguistics*, pp. 131-183. London: Frances Pinter.

Fazel, S. and Grann, M. (2006) 'The population impact of severe mental illness on violent crime', *American Journal of Psychiatry* 163(8): 1397-1403.

File, K. (2018) '"You're a Manchester United manager, you can't say things like that": impression management and identity performance by professional football managers in the media', *Journal of Pragmatics* 127: 56-70.

Firth, J. (1957) 'A synopsis of linguistic theory 1930-1955', in Palmer, E. (eds) *Studies in Linguistic Analysis*, Philological Society, Oxford; reprinted in Palmer, E. (ed.) (1968) *Selected Papers of J. R. Firth*. London: Longman.

Firth, J. R. (1957) *Papers in Linguistics 1934–1951*. Oxford: Oxford University Press.

Flowerdew, J. (2002) 'Globalization discourse: A view from the East', *Discourse and Society* 13(2): 209-225.

Flynn, S., Gask, L. and Shaw, J. (2015) 'Newspaper reporting of homicide-suicide and mental illness', *BJPsych Bulletin* 39(6): 268-272.

Flynn, S., Gask, L., Appleby, L. and Shaw, J. (2016) 'Homicide–suicide and the role of mental disorder: A national consecutive case series', *Social Psychiatry and Psychiatric Epidemiology* 51(6): 877-884.

Foucault M (1972) *The Archaeology of Knowledge and The Discourse on Language*. Pantheon Books, New York

Fowler, R. (1977) *Linguistics and the Novel*. London: Methuen.

Fowler, R. (1991) *Language in the News: Discourse and Ideology in the Press.* London: Routledge

Fowler, R. and Kress, G. (1979) 'Critical linguistic*s'*, in Fowler, R., Hodge, B. Kress, G. and Trew, T. (eds*) Language and Control*, pp. 184-213. London: Routledge.

Fowler, R., Hodge, B., Kress, G. and Trew, T. (1979) *Language and control.* London: Routledge.

Francis, C., Pirkis, J., Francis, C., Pirkis, J., Blood, R. W., Dunt, D. and Putnis, P. (2004) 'The portrayal of mental health and illness in Australian non-fiction media', *Australian and New Zealand Journal of Psychiatry* 38(7): 541-546.

Fung, K. M., Tsang, H. W. and Cheung, W. M. (2011) 'Randomized controlled trial of the self-stigma reduction program among individuals with schizophrenia', *Psychiatry Research* 189(2): 208-214.

Furnham, A. and Baguma, P. (1999) 'Cross-cultural differences in explanations for health and illness: A British and Ugandan comparison', *Mental Health, Religion and Culture* 2(2): 121-134.

Gabrielatos, C. (2007) 'Selecting query terms to build a specialised corpus from a restricted-access database', *ICAME Journal* 31: 5-44.

Gabrielatos, C. and Baker, P. (2008) 'Fleeing, sneaking, flooding: A corpus analysis of discursive constructions of refugees and asylum seekers in the UK press, 1996-2005', *Journal of English Linguistics* 36(1): 5-38.

Galasinski, D. (2017) *Discourses of Men's Suicide Notes: A Qualitative Analysis*. London: Bloomsbury Publishing.

Garside, R. (1987) 'The CLAWS word-tagging system', in Garside, R., Leech, G. and Sampson, G. (eds) *The Computational Analysis of English: A Corpus-based Approach*, pp. 30-41. London: Longman.

Georgaca, E. (2013) 'Social constructionist contributions to critiques of psychiatric diagnosis and classification', *Feminism and Psychology* 23(1): 56-62.

Gernsbacher, M. A. (2017) 'Editorial perspective: The use of person-first language in scholarly writing may accentuate stigma', *Journal of Child Psychology and Psychiatry* 58(7), 859-861.

Gibson, G. (2003) '*Sun* on the ropes over 'bonkers Bruno' story', [online]. Available at: https://www.theguardian.com/media/2003/sep/23/pressandpublishing.mentalhealth [Accessed 12 March 2019].

Goffman, E. (1974) *Stigma: Notes on the Managment of Spoiled Identity*. New York: Penguin Books.

Goulden, R., Corker, E., Evans-Lacko, S., Rose, D., Thornicroft, G. and Henderson, C. (2011) 'Newspaper coverage of mental illness in the UK, 1992-2008', *BMC Public Health* 11(1): 796.

Granello, D. H. and Gibbs, T. A. (2016) 'The power of language and labels: "the mentally ill" versus "people with mental illnesses"', *Journal of Counseling and Development* 94(1): 31-40.

Gray, H. L., Jimenez, D. E., Cucciare, M. A., Tong, H. Q. and Gallagher-Thompson, D. (2009) 'Ethnic differences in beliefs regarding Alzheimer disease among dementia family caregivers', *The American Journal of Geriatric Psychiatry* 17(11): 925-933.

Grice, H. P. (1975) 'Logic and conversation' in Cole, P. and Morgan, J. (eds) *Studies in Syntax and Semantics III: Speech Acts*, pp.183-98. New York: Academic Press.

Gries, S. T. (2005) 'Null hypothesis significance testing of word frequencies: a follow-up on Kilgarriff', *Corpus Linguistics and Linguistic Theory* 1: 277-94.

Gries, S. T. (2010) Corpus linguistics and theoretical linguistics: A love–hate relationship? Not necessarily…', *International Journal of Corpus Linguistics*, 15(3): 327-343.

Gries, S. T. (2013) '50-something years of work on collocations', *International Journal of Corpus Linguistics* 18(1): 137-166.

Grieve, J. (2007) 'Quantitative authorship attribution: an evaluation of techniques', *Literary and Linguistic Computing* 22(3): 251-270.

Grilo, C. M., Sinha, R. and O Malley, S. S. (2002) 'Eating disorders and alcohol use disorders', *Alcohol Research and Health* 26(2): 151-157.

Grondelaers, S. and Geeraerts, D. (1998) 'Vagueness as a euphemistic strategy. *Speaking of Emotions: Conceptualisation and Expression'*, in Athanasiadou, A. and Tabakowska, E. (eds) *Speaking of Emotions: Conceptualisation and Expression,* pp. 357-374. Berlin: Mouton de Gruyter.

Grundmann, R. and Krishnamurthy, R. (201) 'The discourse of climate change: a corpus-based approach', *Critical Approaches to Discourse Analysis Across Disciplines*, 4(2): 125-146.

Guarniero, B., Bellinghini, F. and Gattaz, W. F. (2017) 'The schizophrenia stigma and mass media: a search for news published by wide circulation media in Brazil', *International Review of Psychiatry* 29(3): 241-247.

Gwarjanski, A. R. and Parrott, S. (2018) 'Schizophrenia in the news: The role of news frames in shaping online reader dialogue about mental illness', *Health Communication* 33(8): 954-961.

Hacking, I. (1999) *The Social Construction of What?* Cambridge, Massachusetts: Harvard University Press.

Haghighat, R. (2008) 'Schizophrenia as social discourse: How do people use their diagnosis for social action?', *European Psychiatry*, 23(8): 549-560.

Haghighat, R. and Littlewood, R. (1995) 'What should we call patients with schizophrenia? A sociolinguistic analysis', *Psychiatric Bulletin* 19(7): 407-410.

Hallam, A. (2002) 'Media influences on mental health policy: long-term effects of the Clunis and Silcock cases', *International Review of Psychiatry* 14(1): 26-33.

Halliday, M. A. K. (1985) 'Systemic Background-Systemic Perspectives on Discourse' in Benson, J., D. and Greaves, W., S. (eds) *Selected Theoretical*

Papers from the 19th International Systemic Workshop, pp. 1-15. Norwood, NJ: Ablex Publishing Co.

Halliday, M. A. K. (1985) *An Introduction to Functional Grammar*. London: Edward Arnold.

Halliday, M. A. K. (2002) *On Grammar in the Collected Works of M.A.K. Halliday*. London: Continuum.

Halliday, M. A. K. (2003) 'On the "architecture" of human language' in Webster, J. J. (ed.) *On Language and Linguistics: M. A. K. Halliday*, pp. 1- 29. London: Continuum.

Halliday, M. A. K. (2003 [1990]) 'New ways of meaning: the challenge to applied linguistics' in Webster, J. J. (ed.) *On Language and Linguistics: M. A. K. Halliday*, pp. 139- 174. London: Continuum

Halliday, M. A. K. and Hasan, R. (1976) *Cohesion in English*. London: Longman.

Halliday, M. A. K. and Matthiessen, C. (2004) *An Introduction to Functional Grammar*. *London: Hodder*.

Halliday, M.A.K. (2003 [1973]) *Explorations in the Functions of Language*. London: Edward Arnold.

Halmari, H. (2011) 'Political correctness, euphemism, and language change: The case of 'people first'', *Journal of Pragmatics*, 43(3): 828-840.

Hardie, A. and McEnery, T. (2010) 'On two traditions in corpus linguistics, and what they have in common', *International Journal of Corpus Linguistics* 15(3): 384-394.

Harvey, K. (2012) 'Disclosures of depression: Using corpus linguistics methods to examine young people's online health concerns', *International Journal of Corpus Linguistics* 17(3): 349-379.

Harvey, K. (2013) *Investigating adolescent Health Communication: A Corpus Linguistics Approach*. London: Bloomsbury Publishing.

Harvey, K. (2013) 'Medicalisation, pharmaceutical promotion and the Internet: a critical multimodal discourse analysis of hair loss websites', *Social Semiotics* 23(5): 691-714.

Harvey, K. and Brown, B. (2012) 'Health communication and psychological distress: Exploring the language of self-harm', *Canadian Modern Language Review* 68 (3): 316-340.

Hasan, R. (1987) 'The grammarian's dream: Lexis as most delicate grammar', *New developments in systemic linguistics* 1: 184-211.

Hasson, D.and Arnetz, B. B. (2005) 'Validation and findings com- paring VAS vs. Likert scales for psychosocial measurements', *International Electronic Journal of Health Education* 8: 178–192.

Hodder, I. (2000 [2003]) 'The interpretation of documents and material culture', in Denzin, N. K. and Lincoln, Y. S. (eds) *Handbook of Qualitative Research* (2nd ed.), pp. 703-715. Thousand Oaks, CA: SAGE.

Hoover, D. (2001) 'Statistical stylistics and authorship attribution: an empirical investigation', *Literary and Linguistic Computing* 16(4): 421–444.

Howe, L., Tickle, A. and Brown, I., (2014) ''Schizophrenia is a dirty word': service users' experiences of receiving a diagnosis of schizophrenia', *The Psychiatric Bulletin*. 38(4): 154–158.

Hughes, J. (2008) 'The psychological validity of collocation: evidence from event-related brain potentials', Unpublished PhD thesis. Lancaster University.

Hunston, S. (2001) 'Colligation, lexis, pattern, and text' in Hoey, M., Scott, M. and Thompson, G. (eds) *Patterns of Text: In Honour of Michael Hoey,* pp. 13-33. Amsterdam: John Benjamins.

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

Hunston, S. (2007) 'Semantic prosody revisited', *International Journal of Corpus Linguistics* 12(2): 249-268.

Hunt, D. and Harvey, K. (2015) 'Health Communication and Corpus Linguistics: Using Corpus Tools to Analyse Eating Disorder Discourse Online', in Baker, P. and McEnery, T. (eds) *Corpora and Discourse Studies*. Basingstoke: Palgrave.

Jeffries, L. (2010) *Critical Stylistics: The Power of English*. Basingstoke: Palgrave Macmillan.

Jeffries, L. and McIntyre, D. (2010) *Stylistics*. Cambridge: Cambridge University Press.

Johnson, A. and Woolls, D. (2009) 'Who wrote this? The linguist as detective' in Hunston, S. and Oakey, D. (eds) *Doing Applied Linguistics*, pp. 111-118. London: Routledge

Johnson, S., Culpeper, J. and Suhr, S. (2003) 'From politically correct councillors' to Blairite nonsense': discourses of political correctness' in three British newspapers', *Discourse and Society*, 14(1): 29-47.

Jones, L., Mills, S., Paterson, L. L., Turner, G. and Coffey-Glover, L. (2017) Identity and naming practices in British marriage and civil partnerships. *Gender and Language* 11(3): 309–335

Kenez, S., O'Halloran, P. and Liamputtong, P. (2015), 'The portrayal of mental health in Australian daily newspapers', *Australian and New Zealand Journal of Public Health* 39(6),: 513-517.

Kilgarriff, A. (2005) 'Language is never, ever, ever random', *Corpus Linguistics and Linguistic Theory* 1: 263-75

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P. and Suchomel. V. (2014) 'The Sketch Engine: ten years on', *Lexicography* 1: 7-36.

Kilgarriff, A., Pavel, R., Smrz, P. and Tugwell, D. (2004) 'The Sketch Engine', *Proceedings of Euralex*, 1: 105–116.

Kinloch, K. and Jaworska, S. (2019) 'Using a comparative corpus-assisted approach to study health and illness discourses across domains: the case of postnatal depression (PND) in lay, medical and media texts.' in

Demjen, Z. (ed.) *Contemporary Applied Linguistics in Illness and Healthcare Contexts*. London: Bloomsbury.

Klin, A.and Lemish, D. (2008) 'Mental disorders stigma in the media: review of studies on production, content, and influences', *Journal of Health Communication* 13(5): 434-449.

Knapton, O. (2013) 'Pro-anorexia: Extensions of ingrained concepts', *Discourse and Society* 4(4): 461-477.

Knight, D. (2011) 'The future of multimodal corpora', *Revista brasileira de linguística aplicada*, 11(2): 391-415.

Koike, S., Yamaguchi, S., Ojio, Y., Ohta, K. and Ando, S. (2016) 'Effect of name change of schizophrenia on mass media between 1985 and 2013 in Japan: a text data mining analysis', *Schizophrenia Bulletin* 42(3): 552-559.

Koller, V. and Mautner, G. (2004) 'Computer applications in critical discourse analysis' in Coffin, C., Hewings, A. and O'Halloran, K. (eds) *Applying English Grammar: Functional and Corpus Approaches,* pp. 216-228. London: Hodder and Stoughton.

Koller, V., Hardie, A., Rayson, P. and Semino, E. (2008) 'Using a semantic annotation tool for the analysis of metaphor in discourse', *Metaphorik.de* 15(1): 141-160.

Koteyko, N. and Atanasova, D. (2018) 'Mental health advocacy on Twitter: Positioning in Depression Awareness Week tweets', *Discourse, Context and Media* 25: 52-59.

Kress, G. and Hodge, R. (1979) *Language as Ideology*. London: Routledge.

Kvaale, E. P., Haslam, N. and Gottdiener, W. H. (2013) 'The 'side effects' of medicalization: A meta-analytic review of how biogenetic explanations affect stigma. *Clinical Psychology Review* 33(6): 782-794.

Lam, C.S., Angell, B., Tsang, H.W.H., Shi, K., Corrigan, P.W., Jin, S., Lee, Y.T. and Larson, J.E. (2010) 'Chinese lay theory and mental illness stigma: implications for research and practices', *Journal of Rehabilitation* 76: 35–40.

Lampropoulos, D., Wolman, A. and Apostolidis, T. (2017) 'Analyzing the presentation and the stigma of schizophrenia in French newspapers', *Social Psychiatry and Psychiatric Epidemiology* 52(12): 541-1547.

Lasalvia, A., Penta, E., Sartorius, N. and Henderson, S. (2015) 'Should the label" schizophrenia" be abandoned?', *Schizophrenia Research* 162(1-3): 276-284.

Levinson, S. (1983) *Pragmatics*. Cambridge: Cambridge University Press

Levinson, S. (2000) *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge, Massachusetts: MIT Press.

Lirola, M. M. and Chovanec, J. (2012) 'The dream of a perfect body come true: multimodality in cosmetic surgery advertising', *Discourse & Society 23*(5): 487-507.

Louw, B. (1993) 'Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies', in Baker, M., Francis, G. and Tognini-

Bonelli, E. (eds) *Text and Technology: In Honour of John Sinclair*, pp. 157–76. Amsterdam: John Benjamins.

Louw, B. (2000) 'Contextual Prosodic Theory: bringing semantic prosodies to life', in. Heffer, C. and Sauntson, H. (eds) *Words in Context*. *In Honour of John Sinclair, pp.* 48–94. Birmingham: ELR

Lum, C. C. and Ellis, A. W. (1994) 'Is" nonpropositional" speech preserved in aphasia?', *Brain and Language* 46(3): 368-391.

Lyons, M. and Ziviani, J. (1995) 'Stereotypes, stigma, and mental illness: Learning from fieldwork experiences', *The American Journal of Occupational Therapy* 49(10): 1002-1008.

Machlin, A., King, K., Spittal, M. and Pirkis, J. (2014) 'The role of the media in encouraging men to seek help for depression or anxiety', *Melbourne: Centre for Mental Health, Melbourne School of Population and Global Health*.

Magliano, L., Read, J. and Marassi, R. (2011), 'Metaphoric and non-metaphoric use of the term "schizophrenia" in Italian newspapers', *Social Psychiatry and Psychiatric Epidemiology* 46(10): 1019-1025.

Malinowski, B. (1923) 'The problem of meaning in primitive languages', in Ogden, C. K. and Richards, I., A. (eds) *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism*., pp. 296-336. London: Kegan Paul.

Mann, C. E. and Himelein, M. J. (2004) 'Factors associated with stigmatization of persons with mental illness', *Psychiatric Services* 55(2): 185-187.

Marra, M. (2012) 'Disagreeing without being disagreeable: Negotiating workplace communities as an outsider', *Journal of Pragmatics* 44(12): 1580-1590.

Martin, J. R. (2016) 'Meaning matters: a short history of systemic functional linguistics', *Word* 62(1); 35-58.

Marzano, L., Fraser, L., Scally, M., Farley, S. and Hawton, K. (2018) 'News coverage of suicidal behavior in the United Kingdom and the Republic of Ireland', *Crisis* 39: 386: 396.

Matthiessen, C. (2012) 'Systemic-Functional linguistics as applicable linguistics: social accountability and critical approaches', *D.E.L.T.A.,* 28: 435-471.

Matthiessen, C. M. and Halliday, M. A. K. (1997) *Systemic Functional Grammar: A First Step into the Theory*. Beijing: Higher Education Press.

McArthur, T. and McArthur, T. G. (1981) *Longman lexicon of contemporary English*. London: Longman.

McCabe, R. (2007) 'Conversation analysis', in Slade, M. and Priebe. S. (eds) *Choosing methods in mental health research,* pp. 40-62. Hove: Routledge.

McEnery, T. and Hardie, A. (2012) *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.

McGinty, E. E., Kennedy-Hendricks, A., Choksy, S. and Barry, C. L. (2016) 'Trends in news media coverage of mental illness in the United States: 1995–2014', *Health Affairs* 35(6): 1121-1129.

McIntyre, D. (2018) 'Irony and semantic prosody revisited', in Jobert, M. and Sorlin, S. (eds) *The Pragmatics of Irony and Banter*, pp 81-99. Amsterdam: John Benjamins.

McIntyre, D. and Walker, B. (2019) *Corpus Stylistics: Theory and Practice*. Edinburgh: Edinburgh University Press.

McIntyre, D., Moores, Z., Price, H. and Vice, J. (2018) 'Subtitling Parliament', Paper presented at PALA 2018: Styles and Methods, University of Birmingham, 25-28 July.

McNally, R. J. (2009) 'Can we fix PTSD in DSM-V?' *Depression and Anxiety* 26(7): 597-600.

Mead, G. H. (1934) *Mind, Self and Society*. University of Chicago Press.: Chicago.

Mental Health Act 1983, Chapter 20. (2017) [online]. Available at: http://www.legislation.gov.uk/ukpga/1983/20/pdfs/ukpga_19830020_en.pdf [Accessed: 3 May 2017].

Mental Health Act 2007 (2017) [online]. Available at: http://www.legislation.gov.uk/ukpga/2007/12/pdfs/ukpga_20070012_en.pdf [Accessed: 3 May 2017].

Micali, N., Hagberg, K. W., Petersen, I. and Treasure, J. L. (2013) 'The incidence of eating disorders in the UK in 2000–2009: findings from the General Practice Research Database', *BMJ Open* 3(5)

Mills, S. (2003) 'Caught between sexism, anti-sexism and political correctness': feminist women's negotiations with naming practices', *Discourse and Society* 14(1): 87-110.

Mind (2018) *A-Z Mental Health* [online]. Available at: <https://www.mind.org.uk/information-support/a-z-mental-health/> [Accessed: 3 May 2018]

Mind (2019a) *Mental Health Facts and Statistics*. [online]. Available at: <https://www.mind.org.uk/information-support/types-of-mental-health-problems/statistics-and-facts-about-mental-health/how-common-are-mental-health-problems/#.XIft-i2cbMI> [Accessed 12 March 2019].

Mind (2019b) *What We Do* [online]. Available at: <https://www.mind.org.uk/about-us/what-we-do/> [Accessed 2 Aug 2019].

Mind (2019c) *How to Improve Your Mental Wellbeing* [online]. Available at <https://www.mind.org.uk/information-support/tips-for-everyday-living/wellbeing/#.XYZaxS2ZPMI> [Accessed 5th April 2019].

Mind, (2019d) *Understanding Mental Health Problems* [online] Available at < https://www.mind.org.uk/media/3244655/understanding-mental-health-problems-2016.pdf> [Accessed 5th April 2019].

Mind (2019e) *Post-traumatic Stress Disorder (PTSD) - Symptoms of PTSD.* [online]. Available at: <https://www.mind.org.uk/information-support/types-of-mental-health-problems/post-traumatic-stress-disorder-ptsd/symptoms-of-ptsd/#d> [Accessed 2 Aug 2019].

Mind (2019f) *Dissociation and Dissociative Disorders - Dissociative Disorders* [online]. Available at: <https://www.mind.org.uk/information-support/types-of-mental-health-problems/dissociation-and-dissociative-disorders/dissociative-disorders/?o=6286#.XUnB2i2ZPMJ> [Accessed 2 Aug 2019].

Mind (2019g) *Bipolar Disorder - Bipolar Moods and Symptoms* [online]. Available at: <https://www.mind.org.uk/information-support/types-of-mental-health-problems/bipolar-disorder/bipolar-moods-symptoms/?o=1142#psychosis> [Accessed 2 Aug 2019].

Mind (2019h) *Obsessive-compulsive Disorder (OCD) - about OCD.* [online]. Available at: <https://www.mind.org.uk/information-support/types-of-mental-health-problems/obsessive-compulsive-disorder-ocd/#.XUnL4C2ZPMJ> [Accessed 2 Aug 2019].

Mind (2019i) *Psychosis - Types Of Psychosis*. [online] Available at: <https://www.mind.org.uk/information-support/types-of-mental-health-problems/psychosis/types-of-psychosis/#.XUnMQS2ZPMJ> [Accessed 2 Aug 2019].

Mind (2019j) *Phobias - Types Of Phobia.* [online] Available at: <https://www.mind.org.uk/information-support/types-of-mental-health-problems/phobias/types-of-phobia/#.XUnM8S2ZPMJ> [Accessed 2 Aug 2019].

Mind (2019k) *Depression – Symptoms.* [online] Available at: <https://www.mind.org.uk/information-support/types-of-mental-health-problems/depression/symptoms/?o=9109#.XUnOEC2ZPMJ> [Accessed 2 Aug 2019].

Nairn, R. G. and Coverdale, J. H. (2005) 'People never see us living well: an appraisal of the personal stories about mental illness in a prospective print media sample', *Australian and New Zealand Journal of Psychiatry* 39(4): 281-287.

National Autistic Society (2019) *Mental Health and Autism* [online]. Available at: <https://www.autism.org.uk/about/health/mental-health.aspx> [Accessed 27 March 2019].

Nawka, A., Rukavina, T. V., Nawková, L., Jovanović, N., Brborović, O. and Raboch, J. (2012) 'Psychiatric disorders and aggression in the printed media: is there a link? A central European perspective', *BMC Psychiatry* 12(1): 19.

Nawková, L., Nawka, A., Adámková, T., Rukavina, T. V., Holcnerová, P., Kuzman, M. R. and Miovský, M. (2012) 'The picture of mental health/illness in the printed media in three central European countries', *Journal of Health Communication* 17(1): 22-40.

NHS (2019a) *Schizophrenia.* [online] Available at: <https://www.nhs.uk/conditions/schizophrenia/> [Accessed 2 Aug 2019].

NHS (2019b) *Symptoms - Anorexia Nervosa* [online]. Available at: <https://www.nhs.uk/conditions/anorexia/symptoms/> [Accessed 2 Aug 2019].

NHS (2019c) *Overview: Self Harm*. [online] Available at <https://www.nhs.uk/conditions/self-harm> [Accessed 2 Aug 2019].

O'Halloran, K. (2007) 'Corpus-assisted literary evaluation', *Corpora* 2(1): 33-63.

Ochs, E. and Shohet, M. (2006) 'The cultural structuring of mealtime socialization', *New Directions for Child and Adolescent Development* 2006(111): 35-49.

OED (2019) 'Patient' [online]. Available at <https://www.oed.com/view/Entry/138820?rskey=8ltP1Z&result=1&isAdvanced=false#eid> [Accessed 5th August 2019].

Ohlsson, R. (2018) 'Public discourse on mental health and psychiatry: Representations in Swedish newspapers', *Health 22*(3): 298-314.

Ottewell, N. (2017) 'Newspaper reporting of mental illness', *Journal of Public Mental Health* 16(2): 78-85.

Page, R. (2003) 'Cherie: lawyer, wife, mum': contradictory patterns of representation in media reports of Cherie Booth/Blair', *Discourse and Society*, 14(5): 559-579.

Park, J. H., Choi, Y. M., Kim, B., Lee, D. W. and Gim, M. S. (2012) 'Use of the terms "schizophrenia" and "schizophrenic" in the South Korean news media: a content analysis of newspapers and news programs in the last 10 years', *Psychiatry Investigation* 9(1): 17-24.

Partington, A. (2004) '"Utterly content in each other's company": semantic prosody and semantic preference', *International Journal of Corpus Linguistics* 9(1): 131–156.

Partington, A. (2004) 'Corpora and discourse, a most congruous beast', in Partington, A., Morley, J. and Haarman, L. (eds) *Corpora and Discourse*, pp. 11–20. Bern: Peter Lang.

Partington, A. (2006) 'Metaphors, motifs and similes across discourse types: Corpus assisted Discourse Studies (CADS) at Work', in Stefanowitsch, A. and Gries, S. (eds) *Corpus-based Approaches to Metaphor and Metonymy*, pp. 267–304. Berlin: Mouton de Gruyter.

Partington, A., Duguid, A. and Taylor, C. (2013) *Patterns and Meanings in Discourse: Theory and Practice in Corpus-Assisted Discourse Studies (CADS)* Amsterdam: John Benjamins.

Paterson, B. (2006) 'Newspaper representations of mental illness and the impact of the reporting of "events" on social policy: the "framing" of Isabel Schwarz and Jonathan Zito', *Journal of Psychiatric and Mental Health Nursing* 13(3): 294-300.

Peirce C. S. (1906) 'Pragmatism in retrospect: a last formulation', Reprinted in Buchler, J. (ed.) *Philosophical Writings of Peirce*, pp. 269–289. Dover Publications: New York.

Persaud R. (2003) 'Knocking Bruno when he is down', *BMJ: British Medical Journal 327 (*816), Available at <10.1136/bmj.327.7418.816.>

Pescosolido, B. A., Monahan, J., Link, B. G., Stueve, A. and Kikuzawa, S. (1999) 'The public's view of the competence, dangerousness, and need for legal coercion of persons with mental health problems', *American Journal of Public Health* 89(9): 1339-1345.

Phillips, M. (1989) *Lexical Structure of Text*. English language research. Birmingham: Birmingham University.

Philo, G. (1997) 'Changing media representations of mental health', *Psychiatric Bulletin* 21(3): 171-172.

Philo, G., Secker, J., Platt, S., Henderson, L., McLaughlin, G. and Burnside, J. (1994) 'The impact of the mass media on public images of mental illness: media content and audience belief', *Health Education Journal* 53(3): 271-281.

Pingani, L., Sampogna, G., Borghi, G., Nasi, A., Coriani, S., Luciano, M. and Fiorillo, A. (2018) 'How the use of the term "schizo*" has changed in an Italian newspaper from 2001 to 2015: Findings from a descriptive analysis', *Psychiatry Research* 270: 792-800.

PODS (2019) *The Problem of Prevalence – How Common is DID*? [online]. Available at: <https://information.pods-online.org.uk/the-problem-of-prevalence-how-common-is-did/> [Accessed 2 Aug 2019].

Pope, C. W. (2010) 'The bootcamp discourse and beyond', *International Journal of Corpus Linguistics* 15(3): 323-325.

Price, H. (forthcoming) 'How is brand identity linguistically encoded? A case study of BrewDog', *Language and Literature.*

PTSD UK (2019) *Who is Affected by PTSD* [online]. Available at <https://www.ptsduk.org/what-is-ptsd/who-is-affected-by-ptsd/>. [Accessed 2 Aug 2019].

Rampton, B. and Charalambous, C. (2016) 'Breaking classroom silences: A view from linguistic ethnography', *Language and Intercultural Communication* 16(1): 4-21.

Rayson, P. (2009) Wmatrix: a web-based corpus processing environment, Computing Department, Lancaster University. http://ucrel.lancs.ac.uk/wmatrix/

Rayson, P. and Garside, R. (2000) 'Comparing corpora using frequency profiling', *Proceedings of the Workshop on Comparing Corpora*, pp. 1-6. 38th Annual Meeting of the Association for Computational Linguistics. 1-8 October. Hong Kong.

Read, J. and Law, A. (1999) 'The relationship of causal beliefs and contact with users of mental health services to attitudes to the 'mentally ill'', *International Journal of Social Psychiatry* 45(3): 216-229.

Read, J., Haslam, N., Sayce, L. and Davies, E. (2006) 'Prejudice and schizophrenia: a review of the 'mental illness is an illness like any other' approach', *Acta Psychiatrica Scandinavica* 114(5): 303-318.

Rhydderch, D., Krooupa, A. M., Shefer, G., Goulden, R., Williams, P., Thornicroft, A. and Henderson, C. (2016) 'Changes in newspaper coverage of mental illness from 2008 to 2014 in England', *Acta Psychiatrica Scandinavica* 134(446*)*: 45-52.

Richardson J. (2007) *Analysing Newspapers: An Approach from Critical Discourse Analysis.* Basingstoke: Palgrave Macmillan.

Roberts, E., Bourne, R. and Basden, S. (2013) 'The representation of mental illness in Bermudian print media, 1991–2011', *Psychiatric Services* 64(4): 388-391.

Rogers, R. (ed.) (2004) *An Introduction to Critical Discourse Analysis in Education*. Abingdon: Routledge.

Rosch, E. (1975) 'Cognitive representations of semantic categories', *Journal of Experimental Psychology* 104(3): 192-233.

Rose, D. (1998) 'Television, madness and community care', *Journal of Community and Applied Social Psychology* 8(3): 213-228.

Rowe, R., Tilbury, F., Rapley, M. and O'Ferrall, I. (2003) '"About a year before the breakdown I was having symptoms": sadness, pathology and the Australian newspaper media', *Sociology of Health and Illness* 25(6): 680-696.

Ryan, E. B., Bannister, K. A. and Anas, A. P. (2009) 'The dementia narrative: Writing to reclaim social identity', *Journal of Aging Studies* 23(3): 145-157.

Sagi, E., Kaufmann, S. and Clark, B. (2009) 'Semantic density analysis: Comparing word meaning across time and phonetic space' in *Proceedings of the Workshop on Geometrical Models of Natural Language Semantics* pp. 104-111. Association for Computational Linguistics.

Sampogna, G., Bakolis, I., Robinson, E., Corker, E., Pinfold, V., Thornicroft, G. and Henderson, C. (2017) 'Experience of the Time to Change programme in England as predictor of mental health service users' stigma coping strategies' *Epidemiology and Psychiatric Sciences* 26(5): 517-525.

Sapir, E. (1921) *An Introduction to the Study of Speech*. New York: Harcourt, Brace.

Sayakhot, P., Vincent, A. and Teede, H. (2012) 'Breast cancer and menopause: perceptions of diagnosis, menopausal therapies and health behaviors', *Climacteric* 15(1): 59-67.

Sayer, A. (2000) *Realism and Social Science*. London. Sage.

Scholz, B., Crabb, S. and Wittert, G. A. (2014) '"We've Got to Break Down the Shame" Portrayals of Men's Depression', *Qualitative Health Research*, 24(12): 1648-1657.

Schomerus, G., Auer, C., Rhode, D., Luppa, M., Freyberger, H. J. and Schmidt, S. (2012) 'Personal stigma, problem appraisal and perceived need for

professional help in currently untreated depressed persons', *Journal of Affective Disorders* 139(1): 94-97.

Schomerus, G., Schwahn, C., Holzinger, A., Corrigan, P. W., Grabe, H. J., Carta, M. G. and Angermeyer, M. C. (2012) 'Evolution of public attitudes about mental illness: a systematic review and meta-analysis', *Acta Psychiatrica Scandinavica* 125(6): 440-452.

Scope (2019) 'A social model of disability' [online]. Available at: https://www.scope.org.uk/about-us/social-model-of-disability [Accessed: 20 March 2019].

Scott, M. (1997) 'PC analysis of key words and key key words', *System*, 25(2): 233-245.

Scott, M. (1999) *WordSmith Tools Help Manual, Version 3.0*. Oxford: Mike Scott/Oxford University Press.

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

Sealey, A. and Pak, C. (2018) 'First catch your corpus: methodological challenges in constructing a thematic corpus', *Corpora* 13(2): 229-254.

Searle, J. R. and Willis, S. (1995) *The Construction of Social Reality*. New York: Simon and Schuster.

Semino, E. and Short, M. (2004) *Corpus Stylistics: Speech Writing and Thought Presentation in a Corpus of English Writing*. London: Routledge

Semino, E. (2010) 'Descriptions of pain, metaphor and embodied simulation', *Metaphor and Symbol* 25(4):141-160.

Semino, E. (2011) 'Metaphor, creativity and the experience of pain across genres' in Swann, J., Pope, R. and Carter, R. (eds) *Creativity in Language and Literature: The State of the Art,* pp. 83- 100. Basingstoke: Palgrave Macmillan.

Semino, E., Zakrzewska, J. M. and Williams, A. (2017) 'Images and the dynamics of pain consultations', *The Lancet* 389(10075): 1186-1187.

Short, M. and van Peer, W. (1999) 'A reply to Mackay', *Language and Literature 8*(3): 269-275.

Short, M., Semino, E. and Wynne, M. (2002) 'Revisiting the notion of faithfulness in discourse presentation using a corpus approach', *Language and Literature* 11(4): 325-25

Simmons, L., Jones, T. and Bradley, E. (2017) 'Reducing mental health stigma: the relationship between knowledge and attitude change', *European Journal of Mental Health* 1(12): 25-40.

Simpson, P. (1993) *Language, Ideology and Point of View*. New York; London: Routledge.

Sinclair, J. (1991) *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sinclair, J. (1996) 'The search for units of meaning' *Textus* IX: 75-106.

Sinclair, J. M. (2004) *Trust the Text: Language, Corpus and Discourse*. Abingdon: Routledge.

Slopen, N. B., Watson, A. C., Gracia, G. and Corrigan, P. W. (2007)' Age analysis of newspaper coverage of mental illness', *Journal of health communication* 12(1): 3-15.

Søgaard, A. J. and Fønnebø, V. (1995) 'The Norwegian mental health campaign in 1992. part II: changes in knowledge and attitudes', *Health Education Research* 10(3): 267-278.

Sontag S. (1996) *Illness as Metaphor*. New York: Struss and Giroux.

Stuart, H. (2003) 'Stigma and the daily news: evaluation of a newspaper intervention', *The Canadian Journal of Psychiatry* 48(10): 651-656.

Stubbs, M. (1994) 'Grammar, text, and ideology: computer-assisted methods in the linguistics of representation', *Applied Linguistics* 15(2): 201-223.

Stubbs, M. (1997) 'Whorf's children: Critical comments on critical discourse analysis (CDA)', *British Studies in Applied Linguistics* 12: 100-116.

Stubbs, M. (2001) 'On inference theories and code theories: Corpus evidence for semantic schemas', *Text* 21(3): 437-465.

Stubbs, M. (2001) *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell Publishers.

Tay (2017) 'Using metaphor in healthcare: mental health', in Semino, E. and Demjen, Z. (eds) *The Routledge Handbook of Metaphor and Language,* pp.371-384. Abingdon: Routledge:

Teubert, W. (2005) 'My version of corpus linguistics', *International Journal of Corpus Linguistics*, 10(1): 1-13.

Teubert, W. (2010) 'Our brave new world?', *International Journal of Corpus Linguistics* 15(3): 348-352.

The Samaritans. (2019) *Reporting on Celebrity Deaths*. [online]. Available at: https://www.samaritans.org/about-samaritans/media-guidelines/reporting-celebrity-deaths/> [Accessed 12 March 2019].

Thornicroft, A., Goulden, R., Shefer, G., Rhydderch, D., Rose, D., Williams, P. and Henderson, C. (2013) 'Newspaper coverage of mental illness in England 2008-2011', *The British Journal of Psychiatry* 202(55): 64-69.

Thornton, J. A. and Wahl, O. F. (1996) 'Impact of a newspaper article on attitudes toward mental illness', *Journal of Community Psychology* 24(1): 17-25.

Time to Change (2019) *Mind Your Language!* [online]. Available at: <https://www.time-to-change.org.uk/media-centre/responsible-reporting/mind-your-language> [Accessed 12 March 2019].

Time to change (2019a) *Media Guidelines* [online]. Available at: <https://www.time-to-change.org.uk/sites/default/files/imce_uploads/TtC%20Media%20Leaflet%20NEWS%20(2).pdf> [Accessed 2 Aug 2019].

Time to change (2019b) *Violence and Mental Health* [online]. Available at: <https://www.time-to-change.org.uk/media-centre/responsible-reporting/violence-mental-health-problems>. [Accessed 2 Aug 2019].

Tognini-Bonelli, E. (2001) *Corpus Linguistics at Work*. Amsterdam: John Benjamins.

Torrance, G. W., Feeny, D. and Furlong, W. (2001) 'Visual analog scales: do they have a role in the measurement of preferences for health states?', *Med. Decis. Making* 21: 329-334.

Traugott, E. C. (2010) '(Inter) subjectivity and (inter) subjectification: A reassessment' in Davidse, K., Vandelanotte, L. and Cuyckens, H. (eds) *Subjectification, Intersubjectification and Grammaticalization*, pp29-*71.* Berlin: de Gruyter.

Traugott, E. C. and Dasher, R. B. (2002) *Regularity in Semantic Change*. Cambridge: Cambridge University Press.

Urban, M. (2015) 'Lexical semantic change and semantic reconstruction' in Bowern, C. and Evans, B. (eds) *The Routledge Handbook of Historical Linguistics,* pp. 374-392. Abingdon: Routledge.

Van Dijk, T. A. (1995) 'Aims of critical discourse analysis', *Japanese Discourse* 1: 17-27.

Van Os, J. (2016) '"Schizophrenia" does not exist', *BMJ: British Medical Journal* 352: 375.

Van Peer, W. (1986) *Stylistics and Psychology*. London: Croom Helm.

Vilhauer, R. P. (2015) 'Depictions of auditory verbal hallucinations in news media', *International Journal of Social Psychiatry* 61(1): 58-63.

Wahl, O. F. (2003) 'News media portrayal of mental illness: Implications for public policy', *American Behavioral Scientist* 46(12): 1594-1600.

Wahl, O. F. (2012) 'Stigma as a barrier to recovery from mental illness', *Trends in Cognitive Sciences* 16(1): 9-10.

Wahl, O. F.., Wood, A. and Richards, R. (2002) 'Newspaper coverage of mental illness: is it changing?', *Psychiatric Rehabilitation Skills* 6(1): 9-31.

Walker, B. (2010) 'WMatrix, key concepts and the narrators in Julian Barnes's *Talking It Over*' in McIntyre, D. and Busse, B. (eds) *Language and Style*, pp. 364–387. Basingstoke: Palgrave.

Wallin, A. M. and Ahlström, G. (2010) 'From diagnosis to health: a cross-cultural interview study with immigrants from Somalia', *Scandinavian Journal of Caring Sciences* 24(2): 357-365.

Wang, G. and Chen, Y. (2010) 'Collectivism, relations and Chinese communication', *Chinese Journal of Communication* 3(1): 1-9.

White, P. (2000) 'Media Objectivity and the Rhetoric of News Story Structure' in Ventola, E. (ed.) *Discourse and Community: Doing Functional Linguistics*, pp. 378–97. Tubingen: Gunter Narr Verlag.

Whitley, R.and Wang, J. (2017) 'Good news? A longitudinal analysis of newspaper portrayals of mental illness in Canada 2005 to 2015', *The Canadian Journal of Psychiatry* 62(4): 278-285.

Whitley, R., Adeponle, A. and Miller, A. R. (2015) 'Comparing gendered and generic representations of mental illness in Canadian newspapers: an

exploration of the chivalry hypothesis', *Social Psychiatry and Psychiatric Epidemiology* 50(2): 325-333.

Whitley, R. and Berry, S. (2013) 'Trends in newspaper coverage of mental illness in Canada: 2005–2010', *The Canadian Journal of Psychiatry* 58(2): 107-112.

Whitley, R., Wang, J., Carmichael, V. and Wellen, R. (2017) 'Newspaper articles related to the not criminally responsible on account of mental disorder (NCRMD) designation: a comparative analysis. *Canadian Journal of Psychiatry* 62(10): 697-701.

Whitsitt, S. (2005) 'A critique of the concept of semantic prosody', *International Journal of Corpus Linguistics* 10(3): 283-305.

WHO (2014) *Mental Health: A State of Well-being* [online]. Available at <https://www.who.int/features/factfiles/mental_health/en/> [Accessed 5th April 2018].

WHO (2019) 'Gender and women's mental health' [online]. Available at: <https://www.who.int/mental_health/prevention/genderwomen/en/> [Accessed 20 March 2019].

Whorf, B. L. (1940) 'Science and linguistics', *Technology Review*, 42: 229-231.

Widdowson, H. (1995) 'Discourse Analysis: A Critical View', *Language and Literature* 4 (3): 157-72.

Williams, J., Russell, N. and Irwin, D. (2017) 'On the notion of abstraction in systemic functional linguistics', *Functional Linguistics* 4(1): 13.

Wolfram, W. (2015) 'The Sociolinguistic Construction of African American' in Bloomquist, J., Green, L., J. and Lanehart, S., L. (eds) *The Oxford Handbook of African American Language*, pp. 338-352. Oxford: Oxford University Press.

Wray, A. (2002) *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

Wright, D. (2013) 'Stylistic variation within genre conventions in the Enron email corpus: developing a text-sensitive methodology for authorship research', *International Journal of Speech, Language and the Law* 20(1): 45-75.

Wright, D. and Brookes, G. (2018) '"This is England, speak English!': a corpus-assisted critical study of language ideologies in the right-leaning British press', *Critical Discourse Studies* 16(1):1-28.

Xiao, R.and McEnery, T. (2006) 'Collocation, Semantic Prosody and Near Synonymy: A Cross-Linguistic Perspective', *Applied Linguistics*, 27(1): 103–129.

Yang, Y. and Parrott, S. (2018) 'Schizophrenia in Chinese and US online news media: exploring cultural influence on the mediated portrayal of schizophrenia', *Health Communication* 33(5): 553-561.

Zhang, Y., Jin, Y. and Tang, Y. (2014) 'Framing depression: cultural and organizational influences on coverage of a public health threat and

attribution of responsibilities in Chinese news media, 2000-2012', *Journalism and Mass Communication Quarterly* 92(1): 99-120.

Zhang, Y., Jin, Y., Stewart, S. and Porter, J. (2016) 'Framing responsibility for depression: how US news media attribute causal and problem-solving responsibilities when covering a major public health problem', *Journal of Applied Communication Research* 44(2): 118-135.