Crazy Mad Nutters: The Language of Mental Health

Jena D. Hwang and **Kristy Hollingshead**
Institute for Human and Machine Cognition (IHMC)
Ocala, FL 34470, USA
{jhwang,kseitz}@ihmc.us

Abstract

Many people with mental illnesses face challenges posed by stigma perpetuated by fear and misconception in society at large. This societal stigma against mental health conditions is present in everyday language. In this study we take a set of 14 words with the potential to stigmatize mental health and sample Twitter as an approximation of contemporary discourse. Annotation reveals that these words are used with different senses, from expressive to stigmatizing to clinical.We use these word-sense annotations to extract a set of mental health–aware Twitter users, and compare their language use to that of an age- and gender-matched comparison set of users, discovering a difference in frequency of stigmatizing senses as well as a change in the target of pejorative senses. Such analysis may provide a first step towards a tool with the potential to help everyday people to increase awareness of their own stigmatizing language, and to measure the effectiveness of anti-stigma campaigns to change our discourse.

1 Introduction

The World Health Organization (WHO) estimates that one in four people worldwide will suffer from a mental illness at some point in their lives (World Health Organization, 2011). One in five Americans experience a mental health problem in any given year (Kessler et al., 2007; Substance Abuse and Mental Health Services Administration, 2014). Many people with a mental illness experience social and economic hardship as a direct result of their illness. They must cope with their symptoms, but also with the stigma and discrimination that result from misconceptions about such illnesses (McNair et al., 2002; Corrigan et al., 2003). In fact, the stigma and discrimination related to mental illnesses have been described as having worse consequences than the conditions of the mental illnesses themselves, contributing to people's hesitation to seek treatment for a mental health condition (Corrigan et al., 2014).

Many studies in Linguistics and Cognitive Science have shown that word choice and language use have direct influences on the speaker's thought and actions (c.f., linguistic relativism (Boroditsky, 2011; Berlin and Kay, 1991; Lakoff, 1990)). Word choice and the context to which the words are attributed serve to foster stigma and prejudice toward people with mental health conditions, trivializing serious mental health conditions and their accompanying experiences. Anti-stigma campaigns, designed to raise public awareness of mental stigma, have in recent years focused on bringing public attention to the negative impact of the choice of their words. For example, reporters are advised that, as with any disparaging words related to race and ethnicity, some words should never be used in reporting, including 'crazy', 'nuts', 'lunatic', 'deranged', 'psycho', and 'wacko'. The premise behind anti-stigma campaigns is that increased awareness of the detrimental effects of stigmatizing language associated with mental health will help the public become more judicious in their word choice and, consequently, change their attitudes and behaviors toward mental illnesses and those suffering from them.

One might ask, then, whether these anti-stigma campaigns are effective; are they changing the dis-

Proceedings of the 3rd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, pages 52–62,
San Diego, California, June 16, 2016. ©2016 Association for Computational Linguistics

course, particularly around mental health? Evidence seems to indicate that interventions to reduce stigma are occasionally effective in the short term (Thornicroft et al., 2015). As a first step in addressing this question empirically, we explore 14 of the common terms that have been the focus of a number of anti-stigma campaigns, such as 'crazy', 'mental', or 'psycho', that can be used in a derogatory or pejorative manner. We evaluate if indeed awareness of mental illnesses encourages a more restrained use of these words, either avoiding the words entirely or reducing the use of a word in its stigmatizing sense.

For data, we turn to social media, a platform used by nearly four billion people[1] worldwide. Social media platforms offer an uncensored, unscripted view of the ongoing discourse of the online generation, thus providing a source for analyzing contemporary language use. In particular for this study, we focus on public posts to Twitter.

The paper is structured as follows: we begin with a brief discussion of related work, and present our methods and motivation for gathering social media data from Twitter in a two-stage process. We then inventory the various word senses discovered in the data for each of 14 stigmatizing words related to mental health. We discuss our annotation process, beginning with finer-grained senses and moving to a coarser-grained sense inventory for comparison across the set of stigmatizing words. We show the results of word sense analysis across two different sets of social media users, demonstrating that a user's mental health awareness may be reflected in the use – or lack thereof – of stigmatizing language. Finally, we conclude the paper with a few potential applications of this technology.

2 Related Work

There has recently been an explosion in work using technology to detect and characterize various aspects of individuals with mental health disorders, particularly online (Ayers et al., 2013; Yang et al., 2010; Hausner et al., 2008) and on social media (Coppersmith et al., 2014; De Choudhury, 2013; Nguyen et al., 2014; Schwartz et al., 2013). Many social media users post about their own health conditions, including physical conditions such as cancer or the flu (Paul and Dredze, 2011; Dredze, 2012; Hawn, 2009), but also mental health conditions such as depression (Ramirez-Esparza et al., 2008; De Choudhury et al., 2013; Park et al., 2012), bipolar disorder (Kramer et al., 2004), schizophrenia (Mitchell et al., 2015), and a wide range of other mental health conditions (Coppersmith et al., 2015).

In contrast to this previous work, which analyzed the language use of social media users with mental health conditions, we focus on the use of language related to mental health, regardless of the social media users' mental health status. Reavley and Pilkington (2014) also examined Twitter data related specifically to stigma associated with depression and schizophrenia. Similarly, Joseph et al. (2015) analyzed the sentiment and tone in tweets containing the hashtags #schizophrenia and #schizophrenic, and compared these to tweets containing the hashtags #diabetes and #diabetic, in order to determine the difference in attitude toward an often-stigmatized illness like schizophrenia versus an un-stigmatized physical illness like diabetes. The study discovered that tweets referencing schizophrenia were twice as likely to be negative as tweets referencing diabetes, and that Twitter users were more likely to use humorous or sarcastic language in association with the adjective schizophrenic than with diabetic.

Our work has a broader scope than this previous work, in that we examine stigmatizing language related to a wide range of mental health conditions. The closest work to ours is that of Collins (2015), wherein she conducted a historical topic- and co-occurrence analysis of hashtags #insane, #psycho, '#schizo, and #nutter on Twitter. In this work, we add to this list, as described in the next section (see Table 1), and extend the analysis to include a broader examination of language use.

This work also borrows from annotation methods widely used in natural language processing. The process of annotation includes a linguistic analysis of data, the development of annotation standards or guidelines, and the manual tagging of the data with the set standards. Such annotation methods have been used in annotating data for lexical (Duffield et al., 2010), semantic (Schneider et al., 2015; Hwang et al., 2014; Hwang et al., 2010; Palmer et al., 2005), and syntactic (Marcus et al., 1994) tasks.

[1] http://www.statisticbrain.com/social-networking-statistics/

3 Stigmatizing Words

In this study, we focus on 14 words with the potential to stigmatize mental health, which we will refer to as *stigmatizing words*. Table 1 contains a list of the stigmatizing words used in this study.

bonkers	insane	mad	nuts	schizo
crazy	loony	mental	nutter	wacko
deranged	lunatic	nutcase	psycho	

Table 1: Stigmatizing words used for keyword search.

Starting from the 4 words 'insane', 'psycho', 'schizo', and 'nutter' as studied by Collins (2015), we extended the list by including words that are often cited as problematic terminology by various anti-stigma campaigns to arrive at our list of 14. In particular, we focused on the terminologies discussed in various blog entries, articles, or publications by National Alliance on Mental Illness (NAMI), Time to Change, and HealthyPlace[2].

4 Twitter Data

All of the data in this study comes from publicly available Twitter data collected using the Twitter API. The data was collected in two stages.

4.1 Keyword-Based Data

In the first stage, we obtained two months' (July and August 2015) worth of tweets based on a keyword search of the 14 stigmatizing words. We will refer to the set of tweets collected using this keyword-based search as the *seed set*. This collection contains over 27 million tweets. Once extracted, the seed set was then filtered to remove tweets containing the label RT (retweets) or URLs, on the assumption that such tweets often contain text not authored by the user. Tweets were also filtered to select only those marked by Twitter as English (i.e., "en" or "en-gb"). In order to exclude instances where stigmatizing words show up in the user handle (e.g., @crazygirl), user mentions (@s) were removed for the purposes of filtering. Finally, any exact-match duplicates among the set of tweets were removed for purposes of annotation. The final, filtered set consists of just over 840k tweets. Of these, 'nuts' and 'crazy' make up over 20% of the dataset, each; 'mad' is the next most frequent, at 13%, with 'psycho'; 'insane', and 'mental' following behind with 7-11% each. 'Bonkers' and 'lunatic' each comprise 3% of the data; 'nutter' and 'deranged' are less than 2% of the data, while the remaining words – 'loony', 'nutcase', 'schizo', and 'wacko' – each comprise less than 1% of the filtered seed set.

For each stigmatizing word, 100 random tweets containing the word were selected for annotation. Each selected tweet was manually analyzed to establish an inventory of types and varieties of meanings or *senses* of the words as used in the tweets (see Section 5 for further discussion). Based on the established senses, whenever a tweet was used in a clinical sense of the word, the tweet was considered to originate from a *mental health aware* (MHA) user. Users that did not have any clinical usages were considered to be *mental health unaware* (MHU) users. Table 2 provides a few example tweets for several of the stigmatizing words.

4.2 User-Based Data

The second-stage dataset, which we refer to as the *user-based set*, is constructed from tweets posted publicly by a set of users. For this dataset, we began by using the tweets annotated as MHA to extract a list of Twitter users that had minimally used mental health aware language in at least one tweet. As the tweets typically contained specific references to mental health or mental issues, we make the assumption that these users may be more sensitive to the existing stigma or prejudices towards mental health disorders. We term this set of users the MHA users.

We then extract a set of users for comparison. Generally, a comparison set would be generated based on a random selection of Twitter users. In our case, we limit that selection to users who tweeted any of the stigmatizing words from Table 1 in a non-clinical sense during the collection time period (July-August 2015) and did not use the stigmatizing words in a clinical sense during that time period, and thus do not belong to the MHA user group. We term these users the MHU users. The set of MHU users is much larger than MHA users: 686 versus 60, respectively. We additionally down-selected within the MHU set to form an age- and gender-matched comparison set, based on evidence that failing to ac-

[2]https://www.nami.org/, http://www.time-to-change.org.uk/, and http://www.healthyplace.com/, respectively

"I'm fuming. How dare a TV show portray folks suffering from **mental** health issues so unfairly? As if there isn't already enough stigma around."
"this is exactly why I think JH is borderline personality disorder. seems to fit. no one wanted 2 look at lesser **mental** issue."
"I don't even score high on **schizo** symptoms and that's what bothers me most besides mood issues"
"Ha... apparently according to my **Schizo** voices (audio hallucinations) some of them went out on the piss while I was asleep"
"A lot of Americans are injured for them to portray one person as '**insane** or mentally ill' "
"If you think negatively about yourself, this should help. Im certifiably **nuts**. I know these things."

Table 2: Examples of tweets using stigmatizing words in their clinical sense.

count for age and gender can yield biased comparison groups that may skew results (Dos Reis and Culotta, 2015). To create an approximately matched comparison set, we take each user in our full MHA and MHU sets, and obtain age and gender estimates for each from the tools provided by the World Well-Being Project (Sap et al., 2014). These tools use lexica derived from Facebook data to identify demographics, and have been shown to be successful on Twitter data (Coppersmith et al., 2015). Then, in order to select the best comparison set, we selected (without replacement) the MHU user with the same gender label and closest in age to each of the MHA users.

We use a balanced dataset here for our analysis, by selecting an equal number of MHA users and comparison MHU users for our analysis. In practice and as stated above, we found approximately an order of magnitude difference between MHA and MHU users. Our selection of a balanced set enables simpler machine learning classification efforts, and helps to demonstrate the language differences between the two groups more clearly than if we had examined a dataset more representative of the population ($\sim$1/10). Our results should be taken as validation that the differences in language we observe are relevant to mental health awareness, but only one step towards applying something derived from this research as a tool in a real world scenario.

For each of the MHA users and the age- and gender-matched MHU users, we retrieved all of their most recent public tweets via the Twitter API, up to 3200 tweets.[3] Just as in the previous data collection stage, the data was filtered to select only tweets marked as English, to remove retweets and tweets containing URLs, and to remove user mentions (@s). Preprocessing of the data removed any exact-match duplicates among the filtered tweets.

From this new set of user-based tweets, we extracted each tweet containing any of the stigmatizing words. We then annotated the instances of the stigmatizing words in these extracted tweets with the senses developed in the previous stage (Section 4.1). We achieved a relatively high inter-annotator agreement rates: Cohen's kappa value of 0.74 (IAA = 86.8%) on the MHA tweets and 0.64 (IAA = 80.6%) on MHU tweets.

5 Sense Inventory for Stigmatizing Words

As mentioned in the previous section, the seed set was used to develop an inventory of fine-grained senses for each of the stigmatizing words. Instances annotated with a clinical sense were marked as *MHA tweets* and used to develop a user-based dataset. Subsequently, in order to facilitate a comparison of senses as used across stigmatizing words, the fine-grained senses were binned into coarse-grained senses. This process is detailed below.

5.1 Fine-Grained Sense Inventory

We began with the definitions provided by Merriam-Webster[4] as a basis for our initial sense inven-

[3] 3200 is the maximum number of historic tweets permitted by the API.

[4] http://www.merriam-webster.com/

Sense Definition	Example
Crazy	
1. irrational, crazy	"Maybe I shouldn't be revealing this **crazy** part of me..."
2. excitement	"Got the club going **crazy**!"
3. odd, unusual	"These cigar wraps are **crazy**"
4. extreme	"I miss my best friend like **crazy**."
5. intensifier	"Smh my luck has been **crazy** bad lately"
6. exclamation	"Crazy!"
7. name or label	"Codeine **crazy** goes down in some of the greatest songs ever wrote."
Mad	
1. angry, upset	"@user I'm **mad** at you for being so cute"
2. irrational, crazy	"Keep coughing like a **mad** woman"
3. extreme	"That's a **mad** show"
4. intensifier	"Why is everyone in Chile **mad** good at singing?"
5. exclamation	"**Mad**!" *(akin to "Crazy!")*
6. names or labels	"I just started thinking about a scene from **Mad** Men"
Mental	
1. clinical usage	"**Mental** health awareness is something near and dear to my heart."
2. of the mind	"Emancipate yourselves from **mental** slavery. "
3. irrational, crazy	"People are going **mental** about this lion being killed. "
Nuts	
1. irrational, crazy	"When my mom went **nuts** on my sister for playing hooky..."
2. odd, unusual	"Back to back is **nuts** but meek is about to MURDER it."
3. testicles	"Cassandra showed me her dog's nuts"
4. exclamation	"Aw **nuts** -_-"
5. fruit	"i don't understand why all my office's snacks have **nuts** in them."
6. 'deez nuts'	"if i had a dollar for every time i heard a kid yell "deez **nuts**" at camp i would be rich"
7. clinical use	"Why do you have you Dr's personal cell in your phone?" Uh because I'm **nuts**."
8. building parts	"The **nuts** and bolts for this will be proper implementation and effective evaluation"
Schizo	
1. irrational	"@user Total **schizo** ... I can't imagine using anything other than TweetBot."
2. clinical usage	"I was diagnosed w/addiction once, but turned out I was **schizo**. "
3. names or labels	"I added a video to a playlist **Schizo** BP2635 Brothers Pyrotechnics NEW FOR 2016"

Table 3: Examples of fine-grained sense inventory for the 5 most polysemous words of the 14 stigmatizing words analyzed.

tory. We then made two annotation passes to include missing senses, further refine existing senses, or remove senses as dictated by our annotation of the Twitter data. On average, 4 different senses were identified for the stigmatizing words, ranging from highly polysemous words with 6-8 senses like 'crazy' and 'mad' to words with 1-2 senses like 'deranged' and 'nutcase'. Table 3 shows the sense inventory for the more polysemous words in this study.

As expected, the most common sense for all of the stigmatizing words is the meaning of irrationality or a state of being "not 'right' in one's mind", in reference to a human (e.g., 'crazy' sense 1, 'mad' sense 2, or 'mental' sense 3 in Table 3). Another fairly common meaning includes the state of unusual excitement as attributed to situations (e.g., 'crazy' sense 2 or 'nuts' sense 2) or objects (e.g., 'crazy' sense 3). Note that senses that indicate that someone or something is irrational, extreme, or unusual are considered stigmatizing usages that anti-stigma campaigns highlight.

Additionally, we observe two common functional usages. First is the adverbial usages that function as intensifiers to the adjective the word precedes. For example "crazy bad" in 'crazy' sense 5 and "mad good" in 'mad' sense 4 serve to highlight or reinforce the intensity of the adjectives 'bad' and 'good', respectively. Second common functional use is the

expressive usage as seen for 'crazy' sense 6, 'mad' sense 5, or 'nuts' sense 4. Rather than offering a descriptive content, these expressive serve to convey a certain emotional perspective of the speaker.[5]

Out of the 14 words, only five showed instances that were used in the clinical sense of the word : 'mental' (sense 1), 'nuts' (sense 7), 'psycho' (sense 3), 'insane' (sense 4), and 'schizo' (sense 2). These clinical senses mark the MHA tweets, from which the user-based set was generated as detailed in Section 4.2.

5.2 Coarse-Grained Sense Inventory

If we are to analyze the stigmatizing words and their senses to compare their usage in MHA tweets and MHU tweets, we will find that fine-grained senses pose a difficulty. Consider the senses in Table 3. The words 'schizo' and 'mental', for example, have three senses each, but a sense for one word does not always have a counterpart for the other. Additionally, comparison of usage between highly-polysemous 'nuts' and three-sense 'mental' would require an additional layer of analysis to bridge the differences.

In an effort to make an apples-to-apples comparison of these words, we developed a set of coarse-grained senses that can be applied to all of the stigmatizing words. Fine-grained senses were then mapped to one of five coarse-grained (CGd) senses. Table 4 shows the list of coarse-grained senses.

A.	term applied to sentient beings (often in a derogatory manner)
B.	term applied to an object, situation, or world in general
C.	**clinical usage**
D.	homonymous usage
E.	other senses

Table 4: Coarse-grained (CGd) senses for stigmatizing words.

Sense A, B and C are the relevant senses for our study. Senses A and B capture the stigmatizing sense of the word applied to sentient beings (e.g., humans, pets, etc.) and to objects or situations. Clinical or medical usages of the words are assigned to sense C. To take 'nuts' as an example, sense 1 was mapped to CGd sense A, senses 2 and 4 were mapped to CGd sense B, and sense 7 was mapped to coarse sense C.

Sense D is for homonymous usage like 'mad' sense 1 (i.e. anger in 'I'm mad at you!") or 'nuts' sense 5 (i.e. fruit in "These snacks have nuts in them"). Sense E is a miscellaneous category that includes senses unrelated to the central senses of the word (e.g., a "names and labels" senses) or instances where the sense of the word was not identifiable or unclear (e.g., "tin nuts"). In the case of 'nuts', senses 3 and 5 were mapped to CGd sense D, and senses 6 and 8 were mapped to CGd sense E.

6 MHA vs. MHU

6.1 Stigmatizing Word and Sense Use

In evaluating occurrences of stigmatizing words in MHA and MHU datasets, we find that, on the whole, the MHA users do use these words less frequently when compared to the MHU users. Table 5 shows the total count of tweets in the MHA and MHU datasets for each of the stigmatizing words.[6]

In fact, the number of MHU tweets appears to be nearly double that of MHA tweets. The sheer *lack* of the use of stigmatizing words in MHA suggests that the user's mental health awareness is likely to cause them to be more sensitive towards the stigmatization of those suffering from mental illness. Consequently, they are more likely to shy away from impulsive use of the stigmatizing words.

There are two exceptions to the observation that MHA users use stigmatizing words with less frequency. They are boldfaced in Table 5. The use of the words 'mental' and 'schizo' show higher usage of stigmatized words by the MHA set. However, if we focus on the numbers relevant to the stigmatizing senses (i.e., CGd senses A and B) the story becomes more clear. The leftmost columns of the table show that the majority of uses of the word 'mental' are in fact the clinical sense (CGd sense C). Additionally, note that the use of these words in a clinical sense by MHU users indicates that our original classification of MHU users is imperfect; by our definition, each of the MHU users producing these clinical-sense words should be classified as an MHA user. Since our seed set dataset and our user-based

[5]For example, "Aw nuts!" expresses certain set of emotions (e.g., regret, displeasure) as is relevant to the speaker for the immediate situation, rather than ascribing a descriptive meaning to something or someone.

[6]These counts do not include the homonymous sense D.

	Occurrences		Clinical Use	
	MHA	MHU	MHA	MHU
bonkers	2	19		
crazy	152	252		
deranged	0	12		
insane	24	54	0	1
loony	1	7		
lunatic	0	17		
mad*	22	91		
mental	267	**86**	233	45
nutcase	2	17		
nuts*	7	46	3	2
nutter	1	7		
psycho	17	34	3	0
schizo	12	**9**	9	2
total	489	631	248	50
total (all senses)	610	776		
total (A & B only)	204	503		

Table 5: Word counts of potentially-stigmatizing words in MHA and MHU tweets. An asterisk (*) indicates that the coarse-grained sense D (homonyms) for the word has been removed for this count.

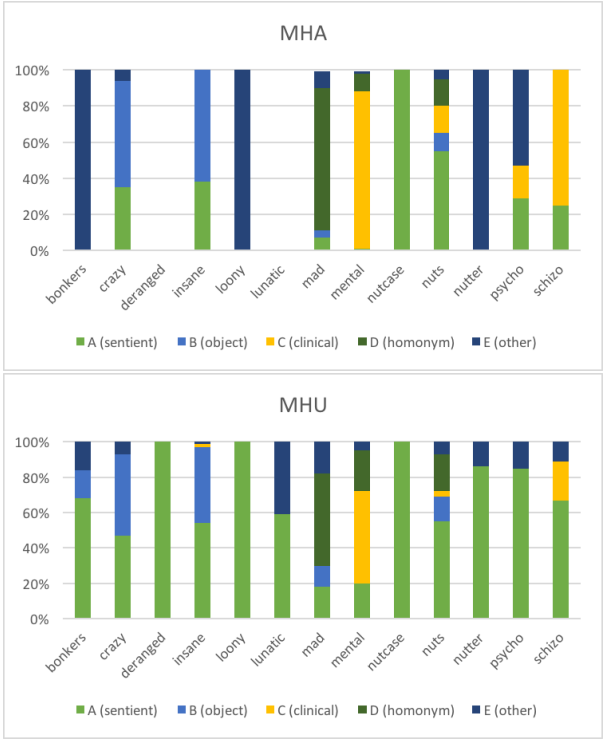

Figure 1: Visualizing coarse-grained senses for stigmatizing words as found in MHA and MHU datasets.

dataset were pulled from Twitter at different time points, approximately six months apart, our classification of MHA and MHU users could be and indeed was occasionally incorrect: MHA users might not have produced another clinical-sense stigmatizing word in the user-based dataset, and MHU users might have used such a sense, perhaps due to becoming more mental health–aware over time.

Figure 1 better visualizes the various senses of each of the stigmatizing words from each of the user groups. Consider the uses of 'mental' and 'schizo' as visualized in Figure 1 for MHA and MHU tweets. For MHA tweets, the most prominent sense of 'mental' is the clinical sense C, while the stigmatizing sense A is used very infrequently. For MHU, however, while there is a large portion of clinical sense C associated with the use of 'mental', it is not as large as that in the MHA set. Sense A also shows up with higher frequency in the MHU set. The same trend can be seen for the word 'schizo'. MHA users do not use stigmatizing sense A as often as the clinical sense C, and the reverse is true for MHU users. As it turns out, although the MHA tweets show a high use of the words 'mental' and 'schizo', most of the usage is attributed to the medical sense. The stigmatizing senses only make up 1% and 25% of the total MHA tweets for 'mental' and 'schizo', respectively, which is considerably lower than the MHU set's stigmatizing sense use, at 20% and 67% for 'mental' and 'schizo', respectively.

Beyond 'mental' and 'schizo', what Figure 1 visually captures is the prominence of the stigmatizing sense A in the MHU group. While sense A does also occur in MHA tweets, the sense is not as prevalent as in the MHU set. The only apparent counter-example in Figure 1 is that of 'nutcase': its only usage is that of sense A in the MHA data. However, note that in Table 5 there were only 2 instances of this word in use, too little data to draw a conclusion.

6.2 Visual Language Analysis

From the previous section, we learned that stigmatizing words do seem to be used differently by MHA and MHU users. In this section, we look to characterize language differences more broadly, by analyzing the set of MHA and MHU tweets as a whole. To do so, we took all of the MHA and MHU tweets gathered for the user-based dataset, extracted the tweets containing any of the 14 stigmatizing words, – all of which had been annotated for their coarse-

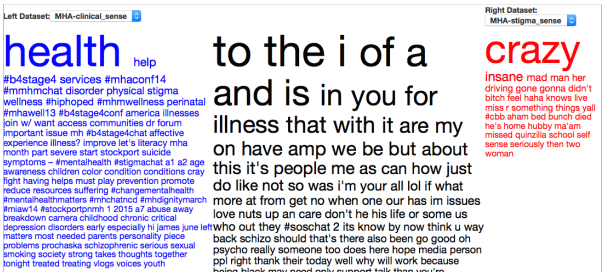

Figure 2: Vennclouds for all tweets containing the 14 stigmatizing words, as tweeted by MHA users, with words from tweets containing clinical senses of the words appearing on the left (blue text), stigmatizing senses of the words appearing on the right (red text), and the language shared by tweets containing either sense in the middle (black text).

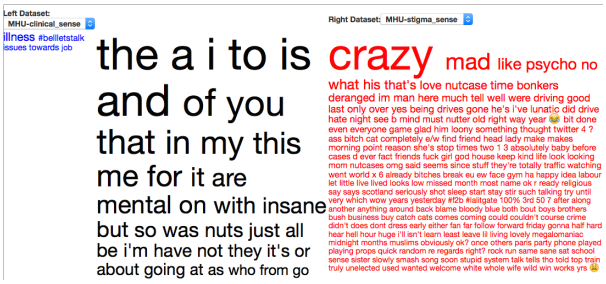

Figure 3: Vennclouds for all tweets containing the 14 stigmatizing words, as tweeted by MHU users, with words from tweets containing clinical senses of the words appearing on the left (blue text), stigmatizing senses of the words appearing on the right (red text), and the language shared by tweets containing either sense in the middle (black text).

grained (CGd) sense – and generated dynamic Vennclouds (Coppersmith and Kelly, 2014) to compare clinical-sense tweets to stigmatizing-sense tweets. Figures 2 and 3 display the resulting clouds.

From these clouds, one might note immediately that in Figure 2, the blue (leftmost) cloud is much larger than the red (rightmost), while the reverse is true in Figure 3. This simply visualizes what was previously quantified in Table 5: MHA users tend to produce more tweets containing clinical sense of the stigmatizing words, whereas MHU users tend to produce more tweets containing stigmatizing senses.

The mere presence of a blue cloud in Figure 3 demonstrates that there was some use of clinical-sense stigmatizing words from the MHU users, who therefore ought to have been categorized as MHA, as previously discussed in Section 6.1. However, on the whole, the simplistic classification of users based on clinical-sense usage held up relatively well.

Both Vennclouds show that 'crazy' was the most frequently-occurring word in stigmatized-sense tweets, shown as the largest, first word in the red (rightmost) clouds of both Figures 2 and 3, with 'mad' as the third- and second-most frequent word, respectively. In fact, these words were never used in a clinical-sense tweet by either of the user groups. This analysis shows quantitatively that, to be more aware of our own uses of stigmatizing senses, we ought to pay particular attention to these words, as the worst offenders from both groups.

The word 'health' is clearly visible as the most frequently-occurring word in the clinical-sense tweets produced by the MHA users, shown as the largest, first word in the blue (leftmost) cloud of Figure 2, with 'help', 'services', 'disorder', and 'stigma' close behind. Perhaps more interesting is the long list of hashtags that appear with high frequency in the clinical-sense tweets from the MHA users, including #b4stage4, #mhaconf14, #mmhmchat, #mhmwellness, #mhawell – all clearly related to mental health, and mental health awareness. This analysis again supports our simplistic classification of users, but additionally gives us a source we could use for another data pull. By simply looking for clinical senses of these stigmatizing words, we discovered clear communities of mental health–aware users.

7 Conclusion & Future Work

In this study we have investigated 14 common terms, used in everyday language, with the potential for stigmatizing mental illnesses in society. We were specifically interested in evaluating if awareness of mental illnesses can help discourage impulsive uses of the pejorative senses of the words. Our findings show that MHA users less frequently use stigmatizing words than MHU users, and when they do use the stigmatizing words, they use the stigmatizing senses less often than their MHU counterparts. Additionally, MHA users tend to structure their language so as to avoid applying the derogatory sense to a sentient being, and use the clinical sense of the stigmatizing word more often than MHU users. The absence of stigmatizing words or, more specifically,

stigmatizing senses by MHA users suggests that the user's mental health awareness contributes to how they employ language in social media, demonstrating a degree of sensitivity towards stigmatization of those with mental illnesses.

Directions for our future work concern the improvement of methods for determining the MHA user group. Our current approach identifies MHA if they show one MHA tweet in the seed data. Unfortunately, basing whether or not a user should is MHA is based on a single tweet does not leave room for the false positive cases, where an otherwise unaware user might have tweeted a aware sounding tweet. Conversely, because the pull for seed set and user-based set were several months apart, there also may have been false negatives – MHU users that should have been classified in the user-based set as MHA users. Most immediate way to address this issue is to experiment by setting a threshold greater than one before a user is considered MHA. We will also look into taking advantage of hashtags related to mental health campaigns to identify users who are intentionally identifying themselves as a part of mental health community. Finally, we intend to also experiment with identifying users that have self-identified as having a mental condition or being a part of mental health community (Coppersmith et al., 2015) as a means of identifying the MHA group.

Another future direction of this work is in further analysis and revision of coarse-grained senses. As discussed in Section 5.1, one of the more frequent fine-grained senses we found are expressives (e.g. "Aww **nuts**!") and intensifiers (e.g. "That's **crazy** good"). These are currently grouped in with the coarse-grained sense B – one of the two stigmatizing senses, but unlike the rest of the fine-grained senses also grouped in B, these usages serve the function of expressing speaker's emotion or emphasizing a descriptive adverbial, rather than carrying a descriptive or content information. In future work, we will look into distinguishing these types of senses from others to determine if indeed these could be considered stigmatizing senses and whether or not these usages are indicative of mental health awareness.

One might imagine several uses for detecting potentially-stigmatizing language. We could use it to warn social media users that their language might been seen as stigmatizing, and offer an opportunity to re-word, similar to the "self-flagging app" mentioned in (Quinn, 2014) to detect potentially offensive or bullying language (Dinakar et al., 2012). This option provides a somewhat heavy-handed method to increase mental health awareness. The ability to automatically detect stigmatizing senses of these (and other) words might also be useful as a filter, to downweight or hide posts containing stigmatizing words, or add a warning or reporting function like Facebook's "offensive content" reporting utility, for stigmatizing language. In this way, users can choose to avoid such language, in case it might trigger negative reactions.

Finally, an automated analysis of the senses in use for potentially-stigmatizing words in everyday language might provide a method to assess whether anti-stigma campaigns are effective. Have we changed the discourse, and if so, in what ways? And, importantly, has a change in discourse resulted in easier access to care, better management of crises, and an improved quality of life for those with mental health conditions? Addressing these questions might shed light on the strengths and weaknesses of the current anti-stigma efforts, and help in guiding future work to end the stigma of mental illness.

Acknowledgments

The authors thank the anonymous reviewers for their thoughtful critiques. Kristy was partially supported in her work by the VA Office of Rural Health Program "Increasing access to pulmonary function testing for rural veterans with ALS through at home testing" (N08-FY15Q1-S1-P01346). The contents represent solely the views of the authors and do not represent the views of the Department of Veterans Affairs or the United States Government.

References

John W. Ayers, Benjamin M. Althouse, Jon-Patrick Allem, J. Niels Rosenquist, and Daniel E. Ford. 2013. Seasonality in seeking mental health information on Google. *American Journal of Preventive Medicine*, 44(5):520–525.

Brent Berlin and Paul Kay. 1991. *Basic color terms: Their universality and evolution*. Univ of California Press.

Lera Boroditsky. 2011. How language shapes thought. *Scientific American*, 304(2):62–65.

Phoebe Collins. 2015. Words will never hurt me; a study of stigmatizing language in Twitter. Via personal communication, 2015-07-12.

Glen Coppersmith and Erin Kelly. 2014. Dynamic wordclouds and Vennclouds for exploratory data analysis. In *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, pages 22–29, Baltimore, Maryland, USA, June. Association for Computational Linguistics.

Glen Coppersmith, Mark Dredze, and Craig Harman. 2014. Quantifying mental health signals in Twitter. In *Proceedings of the ACL Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality (CLPsych)*.

Glen Coppersmith, Mark Dredze, Craig Harman, and Kristy Hollingshead. 2015. From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality (CLPsych)*, pages 1–10, Denver, Colorado, June 5. Association for Computational Linguistics.

Patrick Corrigan, Vetta Thompson, David Lambert, Yvette Sangster, Jeffrey G. Noel, and Jean Campbell. 2003. Perceptions of discrimination among persons with serious mental illness. *Psychiatric Services*.

Patrick W. Corrigan, Benjamin G. Druss, and Deborah A. Perlick. 2014. The impact of mental illness stigma on seeking and participating in mental health care. *Psychological Science in the Public Interest*, 15(2):37–70.

Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. In *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media (ICWSM)*.

Munmun De Choudhury. 2013. Role of social media in tackling challenges in mental health. In *Proceedings of the 2nd International Workshop on Socially-Aware Multimedia*.

Karthik Dinakar, Birago Jones, Catherine Havasi, Henry Lieberman, and Rosalind Picard. 2012. Common sense reasoning for detection, prevention, and mitigation of cyberbullying. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2(3):18.

Virgile Landeiro Dos Reis and Aron Culotta. 2015. Using matched samples to estimate the effects of exercise on mental health from Twitter. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*.

Mark Dredze. 2012. How social media will change public health. *IEEE Intelligent Systems*, 27(4):81–84.

Cecily Jill Duffield, Jena D. Hwang, and Laura A. Michaelis. 2010. Identifying assertions in text and discourse: The presentational relative clause construction. In *Proceedings of Extracting and Using Constructions in Computational Linguistic Workshop held in conjunction wi th NAACL HLT 2010*, Los Angeles, California, June.

Helmut Hausner, Göran Hajak, and Hermann Spießl. 2008. Gender differences in help-seeking behavior on two internet forums for individuals with self-reported depression. *Gender Medicine*, 5(2):181–185.

Carleen Hawn. 2009. Take two aspirin and tweet me in the morning: How Twitter, Facebook, and other social media are reshaping health care. *Health Affairs*, 28(2):361–368.

Jena D. Hwang, Archna Bhatia, Claire Bonial, Aous Mansouri, Ashwini Vaidya, Nianwen Xue, and Martha Palmer. 2010. PropBank annotation of multilingual light verb constructions. In *Proceedings of the Fourth Linguistic Annotation Workshop*, pages 82–90, Uppsala, Sweden, July. Association for Computational Linguistics.

Jena D. Hwang, Annie Zaenen, and Martha Palmer. 2014. Criteria for identifying and annotating caused motion constructions in corpus data. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland.

Adam J. Joseph, Neeraj Tandon, Lawrence H. Yang, Ken Duckworth, John Torous, Larry J. Seidman, and Matcheri S. Keshavan. 2015. #Schizophrenia: Use and misuse on Twitter. *Schizophrenia Research*, 165(2):111–115.

Ronald C. Kessler, Matthias Angermeyer, James C. Anthony, Ron De Graaf, Koen Demyttenaere, Isabelle Gasquet, Giovanni De Girolamo, Semyon Gluzman, Oye Gureje, Josep Maria Haro, Norito Kawakami, Aimee Karam, Daphna Levinson, Maria Elena Medina Mora, Mark A. Oakley Browne, José Posada-Villa, Dan J. Stein, Cheuk Him Adley Tsang, Sergio Aguilar-Gaxiola, Jordi Alonso, Sing Lee, Steven Heeringa, Beth-Ellen Pennell, Patricia Berglund, Michael J. Gruber, Maria Petukhova, and Somnath Chatterji. 2007. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's world mental health survey initiative. *World Psychiatry*, 6(3):168.

Adam D. I. Kramer, Susan R. Fussell, and Leslie D. Setlock. 2004. Text analysis as a tool for analyzing conversation in online support groups. In *Proceedings of the ACM Annual Conference on Human Factors in Computing Systems (CHI)*.

George Lakoff. 1990. *Women, fire, and dangerous things: What categories reveal about the mind*. Cambridge Univ Press.

Mitchell Marcus, Grace Kim, Mary Ann Marcinkiewicz, Robert Macintyre, Ann Bies, Mark Ferguson, Karen Katz, and Britta Schasberger. 1994. The Penn Treebank: Annotating predicate argument structure. In *ARPA Human Language Technology Workshop*, pages 114–119.

Bernard G. McNair, Nicole J. Highet, Ian B. Hickie, and Tracey A. Davenport. 2002. Exploring the perspectives of people whose lives have been affected by depression. *Medical Journal of Australia*, 176(Suppl)(10):S69–S76.

Margaret Mitchell, Kristy Hollingshead, and Glen Coppersmith. 2015. Quantifying the language of schizophrenia in social media. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality (CLPsych)*, pages 11–20, Denver, Colorado, June. Association for Computational Linguistics.

Thin Nguyen, Dinh Phung, Bo Dao, Svetha Venkatesh, and Michael Berk. 2014. Affective and content analysis of online depression communities. *IEEE Transactions on Affective Computing*, 5(3):217–226.

Martha Palmer, Dan Guildea, and Paul Kingsbury. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–105, March.

Minsu Park, Chiyoung Cha, and Meeyoung Cha. 2012. Depressive moods of users portrayed in Twitter. In *Proceedings of the ACM SIGKDD Workshop on Healthcare Informatics (HI-KDD)*.

Michael J. Paul and Mark Dredze. 2011. You are what you tweet: Analyzing Twitter for public health. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM)*.

Cristina Quinn. 2014. MIT algorithm takes aim at social media cyberbullying. http://news.wgbh.org/post/mit-algorithm-takes-aim-social-media-cyberbullying. Accessed 2016-03-01.

Nairan Ramirez-Esparza, Cindy K. Chung, Ewa Kacewicz, and James W. Pennebaker. 2008. The psychology of word use in depression forums in English and in Spanish: Testing two text analytic approaches. In *Proceedings of the 2nd International AAAI Conference on Weblogs and Social Media (ICWSM)*.

Nicola J. Reavley and Pamela D. Pilkington. 2014. Use of Twitter to monitor attitudes toward depression and schizophrenia: an exploratory study. *PeerJ*, 2:e647.

Maarten Sap, Greg Park, Johannes C. Eichstaedt, Margaret L. Kern, David J. Stillwell, Michal Kosinski, Lyle H. Ungar, and H. Andrew Schwartz. 2014. Developing age and gender predictive lexica over social media. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1146–1151.

Nathan Schneider, Vivek Srikumar, Jena D. Hwang, and Martha Palmer. 2015. A hierarchy with, of, and for preposition supersenses. In *Proceedings of the 9th Linguistic Annotation Workshop*, pages 112–123, Denver, Colorado, USA, June. Association for Computational Linguistics.

H. Andrew Schwartz, Johannes C. Eichstaedt, Margaret L. Kern, Lukasz Dziurzynski, Richard E. Lucas, Megha Agrawal, Gregory J. Park, Shrinidhi K. Lakshmikanth, Sneha Jha, Martin E. P. Seligman, and Lyle H. Ungar. 2013. Characterizing geographic variation in well-being using tweets. In *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media (ICWSM)*.

Substance Abuse and Mental Health Services Administration. 2014. Substance abuse and mental health services administration. In *Results from the 2013 National Survey on Drug Use and Health: Mental Health Findings, NSDUH Series H-49, HHS Publication No. (SMA) 14-4887*. Substance Abuse and Mental Health Services Administration, Rockville, MD.

Graham Thornicroft, Nisha Mehta, Sarah Clement, Sara Evans-Lacko, Mary Doherty, Diana Rose, Mirja Koschorke, Rahul Shidhaye, Claire O'Reilly, and Claire Henderson. 2015. Evidence for effective interventions to reduce mental-health-related stigma and discrimination. *The Lancet*.

World Health Organization. 2011. *The World Health Report 2001 – Mental Health: New Understanding, New Hope*. Geneva: World Health Organization.

Albert C. Yang, Norden E. Huang, Chung-Kang Peng, and Shih-Jen Tsai. 2010. Do seasons have an influence on the incidence of depression? The use of an internet search engine query data as a proxy of human affect. *PLOS ONE*, 5(10):e13728.