3 Natural Language Processing for Mental Disorders: An Overview

Iacer Calixto
Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, The Netherlands

Victoria Yaneva
University of Wolverhampton, Wolverhampton, United Kingdom

Raphael Moura Cardoso
University of Brasília, Brasília, Brazil

CONTENTS

DOI: 10.1201/9781003138013-3

3.1 INTRODUCTION

According to the World Health Organization (WHO 2018), mental health refers to "a state of well-being in which the individual realises his/her own abilities, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to the community". In other words, mental health does not just mean the absence of mental disorders. Moreover, the WHO estimates that one in four people using health services suffer from at least one psychiatric or behavioural disorder, and global costs related to mental health are expected to reach US$ 6 trillion by 2030, which is more than double compared to 2010 (Mnookin 2016).

The increasing economic and societal costs of mental disorders worldwide have motivated an increasing body of work on how to use different technologies to assist people living with these conditions. In this chapter, we provide an overview of the Natural Language Processing (NLP) applications and datasets dedicated to addressing problems related to mental health in the NLP literature. We focus on the different applications proposed, the types of data sources these applications use, and the languages they cover.

We first lay out a few crucial definitions we use throughout this chapter. *Psychiatric and behavioural disorders* refer to conditions that negatively affect an individual's mood, cognition or behaviour, and impair social life as well as daily activities. Specifically, *psychiatric disorders* refer to conditions or syndromes that are thought to be the result of some brain dysfunction (e.g., schizophrenia, depressive disorder) and *behavioural disorders* (also known as *psychological disorders*) refer to conditions whose symptoms are primarily observed at the behavioural level (e.g., anorexia, addiction). However, this distinction is blurred because there is no clear-cut evidence of the causal role of the brain or genetic factors for some disorders, which does not mean that such factors do not contribute to the disorder's manifestation (Holmqvist 2013; Borsboom et al. 2019). An example of the difficulty of distinguishing between the two is Autism Spectrum Disorder (ASD), which is considered a developmental disorder of neural origin (American Psychiatric Association 2013), but is still often classified as a behavioural disorder. Since the scientific debate around these definitions and classifications remains outside of the scope of this chapter, we use the term *mental disorders* to refer to any psychiatric or psychological disorders, including neurodevelopmental disorders, and in Section 2.1 we discuss the specific list of conditions we investigate under this umbrella term.

Mental disorders are diagnosed by a psychiatrist or psychotherapist based on a clinical assessment and the patient's self-report (or a third-party report) of specific symptoms and signs. Since this diagnosis is based on behavioural assessment, it is often subjective, may require long periods of observation, and is costly to obtain. These drawbacks have spurred a research line of identifying objective markers of these conditions in behavioural data (e.g., speech, text, or eye-tracking data) to provide additional, objective evidence to the diagnostic process. Another line of research associated with improving the quality of life of people who live with mental disorders is the development of assistive technology such as conversational agents to help people with depression, reading comprehension support tools for people with autism, and spell-checkers that are able to capture dyslexia-specific errors, among others. The development of such diagnostic and assistive tools is a vibrant research area in different technology-related fields, many of which include the processing of language and thus use approaches from the NLP field.

NLP is an inter-disciplinary field applying methodology of computer science and linguistics to the processing of natural languages (Jurafsky and Manning 2012). NLP methods are applicable to clinical assessments written by psychiatrists and psychotherapists (e.g., in electronic health records; EHRs), as well as to utterances or texts produced by patients. Even when not in a clinical or therapeutic setting, e.g., when one writes posts on social media, this information could still be valuable and used to detect *linguistic markers* associated with specific mental disorders, as well as to drive the development of assistive technology.

3.1.1 Main Contributions

Our main contributions in this work are: i) We provide a birds-eye view on the current state of the research being conducted by the NLP research community on topics related to mental disorders with a focus on non-English and low-resource languages; ii) We survey a detailed list of the resources and datasets available for NLP researchers to train and validate data-driven models that address tasks related to mental disorders; and iii) Given the current landscape, we wrap up with suggestions and recommendations on the most important next steps to accelerate and fully realise the potential NLP brings to mental health with special attention to ethical and privacy issues.

3.1.2 Related Surveys

We note that there are a number of related surveys relevant to our topic in this Chapter: Névéol et al. (2018) provide an overview of clinical NLP for languages other than English, where the focus is on *clinical data*, e.g., EHRs; Graham et al. (2019) survey applications of artificial intelligence to mental health, which includes not only NLP but also other fields (e.g., using computer vision to predict mental disorders from brain images); Bzdok and Meyer-Lindenberg (2018) provide an overview on machine learning applied to psychiatric disorders, covering broad *biological markers* (e.g., genetic, behavioural, neural) but not linguistic markers. Finally, Šuster et al. (2017) discuss issues in clinical NLP, such as the difficulty to access high-quality data due to privacy concerns, and highlight different possible sources of bias in data. Our survey differs from previous work in that it draws a landscape of how mental disorders are investigated within the NLP community in particular, with a focus on the different applications, data sources, and languages covered by the NLP literature on mental disorders. Our target audience is NLP researchers and professionals interested in methods, applications, resources and datasets broadly related to mental disorders.

3.2 THE LANDSCAPE OF NLP FOR MENTAL DISORDERS

Our goal in this Section is to provide a high-level, *birds-eye overview* of the area or, in other words, a glimpse into the overall landscape of the research in NLP for mental disorders. To do that, we survey papers related to this topic using the criteria described below. We implement this procedure in software and make our code publicly available[1]. Our search includes papers indexed by the *Association for Computational Linguistics (ACL) Anthology* (Gildea et al. 2018)[2], arguably the database with the most important scientific venues in NLP. This way we provide an overview of how research on mental disorders is approached in top-tier NLP journals, conferences, as well as dedicated workshops.

3.2.1 Inclusion and Exclusion Criteria

We include publications that conform to the following inclusion criteria: 1) Publications must be indexed by the ACL Anthology, i.e., an indexed journal, conference or workshop, without restrictions on the publication year; 2) Publications must include in their title or abstract at least one of the keywords in our keywords list, which we discuss next. Criteria (1) and (2) above are automatic and implemented in software, i.e., one can directly download our code and run it to reproduce the output of criteria (1–2) using the same keywords we use. We now discuss each criterion in more detail.

3.2.1.1 ACL Anthology (criterion 1)

We use an authoritative index to search for the NLP publications to include in this survey. As stated on the ACL public website, the Association for Computational Linguistics “is the premier international scientific and professional society for people working on computational problems involving human language, a field often referred to as either computational linguistics or natural language processing (NLP)”[3]. The ACL Anthology indexes 62, 486 papers across journals, conferences, workshops, and shared tasks, and is arguably the most important index for research publications in NLP.

3.2.1.2 Keyword Matching (criterion 2)

We use the exact substring match between a keyword and the title and abstract, and a paper is considered a match if a keyword is contained in either of the two. We choose keywords to identify papers dealing with mental disorders that have a high estimated burden and are the most epidemiologically prevalent (Whiteford et al. 2015). We use the keywords: “autism”, “ASD”[4], “dyslexia”, “depression”, “obsessive compulsive disorder”, “obsessive-compulsive disorder”, “OCD”[5], “attention deficit and hyperactivity disorder”, “ADHD”[6], “schizophrenia”, “down syndrome”, “substance abuse”, “tobacco abuse”, “tabagism”, “eating disorder”, “anorexia”, “bulimia”, “mental health” and “developmental disorder”.

3.2.1.3 Manual Validation

Next, we manually validate all papers by reading their titles and abstracts and filtering out any incorrectly retrieved papers (false positives), leaving a total of

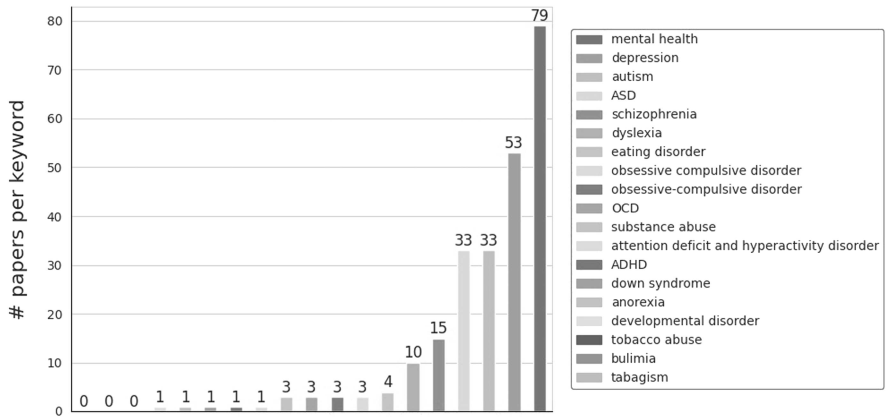

FIGURE 3.1 Number of retrieved papers from the ACL anthology with each keyword. Keywords "tabagism", "bulimia", and "tobacco abuse" retrieved zero papers.

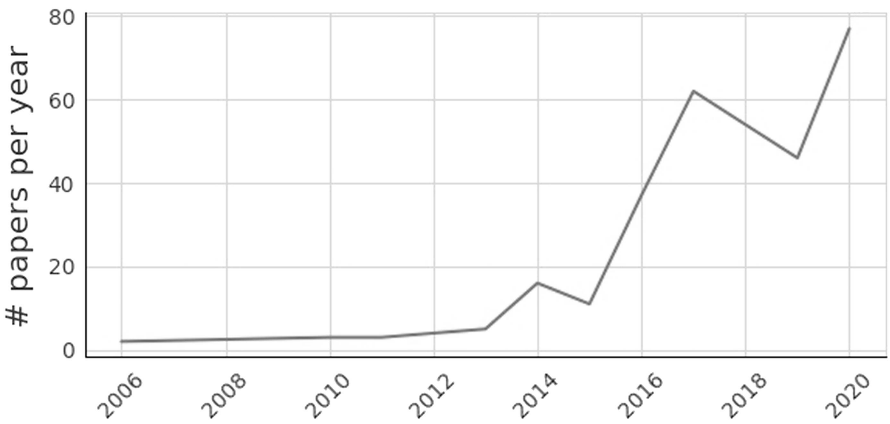

FIGURE 3.2 Number of retrieved papers per year (2006–2020).

Fig 3.1

155 unique publications. In Figure 3.1, we show the number of papers retrieved per keyword and note that the counts shown can include the same paper retrieved with different keywords. For example, there is a large overlap between papers retrieved with keywords "autism" and "ASD", as expected. Overall, the keyword which matches the largest number of papers is "mental health" (79), which in practice is used as an umbrella term for mental health-related work broadly. The keywords "depression", "autism", "ASD" and "schizophrenia" come next, matching a total of 53, 33, 33 and 15 papers, respectively. Conversely, some keywords matched zero publications: "bulimia", "tobacco abuse" and "tabagism".

Fig 3.2

In Figure 3.2, we show the number of papers retrieved per year of publication. We note that there is a clear surge in the number of publications on mental health-related topics in the last five years in NLP venues.

3.2.2 NLP for Mental Disorders: A Birds-Eye View

The trade-off between quality and quantity of data exists in many fields and domains but is especially prominent in the NLP literature related to mental disorders. This can be explained by a combination of multiple factors, such as the difficulty to access participants with verified diagnoses, challenges in data collection further exacerbated by the specifics of certain conditions and, importantly, the ethical implications of making such data publicly available. These restrictions have resulted in some NLP studies using data collected from participants with verified mental disorder diagnoses, where there is high confidence in the data sources but a small amount of data available. One way in which the NLP field has approached this data scarcity problem is through obtaining large amounts of text from social media such as Twitter, Reddit, and others, by scraping content related to mental disorders. In these cases, the abundance of data allows for the application of state-of-the-art NLP techniques but this advantage comes at the high cost of not knowing whether the data comes from participants who indeed have a certain condition, even in cases where the participants themselves think that they do. This foundational distinction between the data sources (verified vs. unverified diagnoses) entails different dataset sizes, which in turn results in different methods, applications, and levels of confidence in the results.

One of the first findings of our review is that the use of these two types of datasets (verified vs. unverified diagnoses) vary across conditions. For example, more narrowly specified conditions such as *autism, dyslexia* or *schizophrenia* are associated with more papers that use data from verified participants, while broader umbrella terms such as *mental health* or *depression* retrieve more papers using social media data. This means that the knowledge gained by many NLP studies in mental health comes from unverified sources, which is a tendency worth noting when discussing the implications of this research.

Below, we outline studies on different mental disorders found in the NLP literature, where we specifically highlight: i) the types of applications, ii) their data sources, and iii) language coverage. In some cases, papers returned under one category (e.g., *mental health* or *developmental disorders*) may, in fact, refer to a more specific category (e.g., *depression* or *autism*), in which case they are discussed under their more specific heading. The same goes for papers that introduce a resource, which may be returned under *mental health* for example, but are discussed in a specific section dedicated to resources (Section 2.3).

3.2.2.1 Autism Spectrum Disorder (ASD)

The majority of papers on ASD use data from verified participants, but datasets are seldomly made publicly available, meaning that there are no replications or reproductions of any of these studies. The main application of interest is the detection of autism at an early age owing to the difficulties of diagnosing the condition (e.g., Tanaka et al. 2014; Parish-Morris et al. 2018; Hauser et al. 2019; Heeman et al. 2010). The majority of these papers focus on the identification of linguistic markers in transcribed speech, often recorded during the diagnostic process, and often in the form of a dialogue between the person with autism and an interlocutor. Other studies that use data from participants with confirmed diagnoses focus on readability research (Yaneva et al. 2017) and complex word identification (Štajner et al. 2017), including a survey on the text adaptation needs of adults with ASD (Yaneva et al. 2019) as well as an analysis of visual-semantic priming and narratives (Regneri et al. 2020). Yaneva et al. (2016) publicly releases a multimodal dataset with text and eye-tracking data collected from adults with and without autism. Finally, among the papers that use data from participants with unverified diagnoses, one study focuses on topic modelling for an ASD support forum (Ji et al. 2014), while other studies focus on text simplification or reading assistance without reporting evaluation with ASD participants (Barbu et al. 2013; Evans et al. 2014). Among all studies on autism, only Tanaka et al. (2014) explores a language that is not English (i.e., Japanese). This results in a serious gap in our knowledge of how linguistic markers of autism are expressed in different languages, whether interlinguistic commonalities exist between these markers, and whether there are common ways to overcome the challenges people with autism may face when reading in different languages.

3.2.2.2 Dyslexia

Unlike the autism studies above, the papers focused on dyslexia explore a greater variety of languages, including Spanish, French, German, Arabic, and Icelandic. Our search for dyslexia returned ten studies, out of which six use data from participants with verified diagnoses. These include three resource papers with annotated writing errors made by dyslexic participants in Spanish (Rello et al. 2014a), German (Rauschenberger et al. 2016) and Arabic (Alamri and Teahan 2017). Other studies focus on the evaluation of reading support tools for dyslexia such as testing whether highlighting the main ideas or including graphical schemes has an impact on readability and comprehensibility for Spanish (Rello et al. 2012, 2014b), or whether lexical simplification helps dyslexic children read French (Gala and Ziegler 2016). The studies that do not rely on participants with confirmed dyslexia are related to automatically creating a set of errors for Icelandic to train a spell checker that is hypothesised to help people with dyslexia (Friðriksdóttir and Ingason 2020), the automatic detection of letter combinations in Shakespeare sonnets which may be found challenging by people with dyslexia (Bleiweiss 2020), and developing a reading tutor for children with dyslexia that is yet to be evaluated (Ward and Crowley 2011). Quiniou and Daille (2018) propose a method to detect constructions in French annotated by a speech therapist that may be challenging for people with dyslexia.

3.2.2.3 Schizophrenia

Similarly to autism, the majority of papers using data from participants with a verified diagnosis of schizophrenia focus on identifying linguistic markers in speech samples that are then used to detect the condition (e.g., Gutiérrez et al. 2017; Just et al. 2019). This is done predominantly for English, with some exceptions being Bulgarian (Boytcheva et al. 2017), French (Amblard et al. 2020), and Hebrew (Bar et al. 2019). Another line of work proposes extracting symptoms from EHRs that could indicate the onset of the condition (Viani et al. 2018; Gorrell et al. 2013). One paper aims to predict the probability of adherence to assigned treatment using dialogues from outpatient consultations (Howes et al. 2012). Finally, several papers mine social media such as Twitter for linguistic markers of schizophrenia based on self-reported diagnosis (e.g., Mitchell et al. 2015; Loveys et al. 2017).

3.2.2.4 Depression

The papers related to depression are the second largest group returned by our search. Among them, 35 studies use data from participants with unverified diagnoses (predominantly from Twitter or Reddit), and 8 use data from participants with verified diagnoses. In the majority of cases, the papers' main goal is to detect the depressive condition (23 papers from the unverified group and six papers with verified participants). This is done through speech analysis (e.g., Lamers et al. 2014) in the cases of papers with verified participants and social media text in the case of papers without verified participants (e.g., Mowery et al. 2016; Song et al. 2018; Husseini Orabi et al. 2018; An et al. 2020). Some studies use both types of data sources, for instances Delahunty et al. (2019). In terms of methods, most papers directly model the condition by approaching its detection as a classification task, with a few offering more knowledge-rich approaches such as semi-automatic emotion annotation (Canales et al. 2016), detecting the use of figurative language in text as an auxiliary task (Yadav et al. 2020), and topic modelling for depression detection (Resnik et al. 2015).

Besides depression detection, other applications include the investigation of patients' attitudes to treatment (Mikal et al. 2017), discourse analysis for support groups (Yip 2018), and quantifying the risk of self-harm (Yates et al. 2017). We note that for some of these applications, e.g., discourse analysis for support groups, the lack of a verified diagnosis is not necessarily a limitation. These studies make a valuable contribution to our understanding of the concerns and coping strategies of people who feel depressed, which can be used as a source of great value in therapy or other social settings. Nevertheless, the lack of a verified diagnosis remains an issue when trying to detect the condition. This is further exacerbated by the fact that many of these studies use a very broad, non-scientific definition of depression, which does not allow verifying the actual usefulness of the tool and its ability to discriminate depression from other conditions, which may have similar linguistic markers (e.g., anxiety, bipolar disorder). Another serious limitation of NLP research related to depression is the lack of linguistic diversity. Studies such as Santos et al. (2020) who study Brazilian Portuguese are a rare exception in this English-dominated sub-domain.

3.2.2.5 General Mental Health

The papers returned for the *mental health* query are the largest category of papers and are fairly homogeneous in terms of data sources in languages. The vast majority use English-language social media data (e.g., Jamil et al. 2017; Mikal et al. 2017; Ireland and Iserman 2018; Loveys et al. 2018) to detect generic mental health concerns, depression or anxiety (often without following a formal definition of these conditions). Some studies such as Gkotsis et al. (2016a) focus on detecting suicide ideation in social media. The few exceptions using data from verified participants include the prediction of adult mental health from childhood writing (Radford et al. 2018), which was part of the 2018 CLPsych shared task, several papers mining electronic health records for information such as suicidality (Gkotsis et al. 2016b) and medication (Chaturvedi et al. 2020). Lamers et al. (2014) are an exception where speech features are used to predict depression using data from therapeutic interventions of (verified) patients living in a nursing home in Amsterdam (i.e., Dutch language). Regarding papers on languages other than English but without verified participants, one uses social media data to predict temporal mental health dynamics (Tabak and Purver 2020), and two propose conversational agents: one for Arabic to be used to help clinicians in administering treatment (Fadhil and AbuRa'ed 2019), and one for Japanese where the goal is to engage senior people using attentive listening techniques (Lala et al. 2017). While the body of work that uses social media to detect mental disorders has obvious limitations entailed by the lack of verified diagnoses, these studies are valuable for identifying new linguistic markers that can, at a later point, be evaluated using controlled datasets. Finally, Amir et al. (2019) use social media data to identify mental health issues "at large" and compare it with prevalence reports.

The applications developed for each of these conditions depend on and are driven by the available resources. To better characterise the relationship between the two, in the next section we present the datasets publicly available in the field of NLP for mental disorders.

3.2.3 Datasets and Resources

This section describes datasets related to mental disorders that are available to the research community. The criteria for including these resources is that the papers describing them specifically state that they are available to other researchers, either through a repository or by contacting the authors and signing a data usage agreement. We do not include papers describing and using resources that were not noted as publicly available.

Table 3.1
Table 3.2

As can be seen in Tables 3.1 and 3.2, the number of resources that use data from participants with verified diagnoses and are publicly available is much smaller than those that use data from social media. Moreover, datasets from participants with verified diagnoses are typically very limited in size (as expected), such as the dataset released by Yaneva et al. (2017) with 27 texts, or Rauschenberger et al. (2016) with 47 texts. Albeit smaller, the resources using data from verified participants are comparatively more diverse in terms of the languages (English, Spanish, German) and conditions (autism, dyslexia, general mental health) they cover; by contrast, the social media resources that are publicly available typically target generic mental health related to anxiety, depression or suicide ideation and focus almost predominantly on English-language sources, notable exceptions being Park et al. (2020) for Korean and Lee et al. (2020) for Cantonese.

Next, in Section 3, among other things, we discuss the implications the availability of these resources have for the development of the field.

3.3 DISCUSSION

We draw three main conclusions based on the review of the papers presented in the previous sections. The first conclusion refers to the fact that the majority of papers in the field, especially the ones which refer to broad terms such as *mental health* and

TABLE 3.1

Summary of resources: Participants whose diagnosis was verified

Paper	Condition	Language	Description	Data Availability
Yaneva et al. (2016)	Autism	English	A multimodal data set of text and eye-tracking was data obtained from autistic and control readers. The data were elicited through reading comprehension tests and includes participants' responses to the questions. An extended version of the data set with 20 text passages and 60 questions is available in Yaneva (2016).	https://github.com/victoria-ianeva/ASD-eye-tracking-reading-corp
Yaneva et al. (2017)	Autism	English	A collection of 27 individual documents whose readability was evaluated by 27 different people diagnosed with autism	https://github.com/victoria-ianeva/ASD-Comprehension-Corpus
Rello et al. (2014a)	Dyslexia	Spanish	54 school essays and homework exercises provided by teachers from children and teenagers with dyslexia between 6 and 15 years old and 29 texts provided by parents with children with dyslexia with 887 misspellt words.	www.luzrello.com/dyslist.html (link not functional, January 22, 2021).
Rauschenberger et al. (2016)	Dyslexia	German	47 texts (homework exercises, dictations, and school essays) written by students from 8 to 17 years old with 1,021 unique errors	http://goo.gl/LRaUDA (link not functional, January 22, 2021).
Lynn et al. (2018)	General mental health	English	Data from the British Birth Cohort Study, which follows a cohort of children born in a single week in March 1958 in Great Britain. The data contains essays written by the children at age 11 and used to predict aspects of their mental health at various ages.	Data was made available as part of the CLPsych 2018 Shared Task.
Gratch et al. (2014)	Anxiety, depression and PTSD	English	Clinical diagnostic interviews conducted by humans, human-controlled agents, and autonomous agents. The participants include both distressed and non-distressed individuals. Data collected include audio and video recordings and extensive questionnaire responses; parts of the corpus have been transcribed and annotated for a variety of verbal and non-verbal features.	"Currently, the corpus is being shared on a case-by-case basis by request and for research purposes".

TABLE 3.2

Summary of resources: Participants with no verified diagnosis

Paper	Condition	Language	Description	Data Availability
Cohan et al. (2018)	General mental health	English	Posts extracted from a publicly available Reddit corpus corresponding to users who claimed to have been diagnosed with a mental disorder (20,406 users) and control users who are unlikely to have one of the mental disorders studied (335,952 users).	Available through a data usage agreement protecting the users' privacy.
Coppersmith et al. (2014)	General mental health	English	Tweets from people who have stated that they have depression (1 million), bipolar disorder (992k), PTSD (573k), seasonal affective disorder (421k), as well as a control group tweets (13.7 million).	Data was made available as part of the CLPsych 2015 Shared Task.
Park et al. (2020)	Suicide risk	Korean	2,791 posts from Korean Q&A site written by active-duty military personnel with 13,955 expert annotations of suicidal risk levels.	Available subject to research ethics agreement.
Anani et al. (2019)	General mental health	English	Shared task description paper that describes the RDoC (Research Domain Criteria framework) task in the BioNLP Open Shared Tasks. The dataset consists of Pubmed abstracts from papers related to mental health.	https://montana.app.box.com/s/kh0hmyn1jcj5ajvr2nibq4iwwgiv3lefolder/73225997826 invalid link May 7, 2021.
Matero and Schwartz (2020)	General mental health	English	A dataset collected from 1,900 Twitter users and annotated for affective language use with weekly and daily scores for six emotions.	https://github.com/MatthewMatero/AffectiveLanguageForecasting

(Continued)

TABLE 3.2 (Continued)

Summary of resources: Participants with no verified diagnosis

Paper	Condition	Language	Description	Data Availability
Yates et al. (2017)	Depression	English	A Reddit dataset consisting of users with self-reported depression diagnoses matched with control users.	"Available to researchers who agree to follow ethical guidelines".
Ghosh et al. (2020)	Suicide	English	Corpus of emotion-annotated suicide notes. The corpus consists of 2,393 sentences from around 205 suicide notes collected from various sources.	https://www.iitp.ac.in/~ai-nlp-ml/resources.html# CEASE
Lee et al. (2020)	General mental health	Cantonese	A corpus designed for training a virtual counsellor, which consists of a domain-independent sub-corpus that supports small talk for rapport building with users, and a domain-specific sub-corpus that provides material for a particular area of counselling.	Available for Research purposes on request to the first author.
Ellendorff et al. (2016)	General mental health	English	175 abstracts of PubMed research articles chosen based on a selection of mental disorders and annotated for events.	http://www.ontogene.org/current-pr/psymine
Milne et al. (2016)	General mental health	English	The ReachOut dataset consists of 65,024 youth-support-forum posts written between July 2012 and June 2015. The posts were annotated in four categories corresponding to levels of distress.	Data was made available as part of the CLPsych 2016 Shared Task.
Zirikly et al. (2019) and Shing et al. (2018)	General mental health	English	1,556,194 Reddit posts from 11,129 users collected from the r/SuicideWatch subreddit and annotated into four categories of suicide risk.	Data was made available as part of the CLPsych 2019 Shared Task.

depression, use data obtained from social media. Second, we note that there is a variety of NLP applications applied to mental disorders, starting with detection of various conditions based on linguistic markers, through the development of assistive technology of various kinds (e.g., conversational agents, reading tutors, text simplification), to topic modelling in support forums and social media. Finally, and somewhat unsurprisingly, the vast majority of the research in the field is done for the English language with very few exceptions focusing on other languages, and almost no studies on under-resourced languages. The paragraphs below discuss each of these findings and their implications for validity, ethics, and equity within the field of NLP for mental disorders.

3.3.1 Types of Data Sources

The fact that the majority of papers use social media data indicates that most of the knowledge in the field is derived from a population sub-sample who uses social media and does not have a verified diagnosis, meaning that there are serious issues related to selection bias and low outcome validity. While there are, undoubtedly, advantages to this approach given the amount of available data, there are also ethical considerations related to informed consent by vulnerable participants. The fact that this material is often public does not exempt researchers from the ethical standards that should guide the research and must be evaluated with extreme caution. For example, there is a potential privacy risk for the non-informed subjects: although data sanitisation is a minimum necessary technique for protecting an individual's privacy, this procedure may not always be sufficient to guarantee anonymity (Malin et al. 2013). To address the potential ethical issues stemming from the use of social media data from vulnerable populations (as is the case of those with mental disorders), we recommend that researchers not only place utmost importance on data anonymisation, but also only share the data subject to research ethics agreement whenever participant-produced language is used.

Having stated the drawbacks and risks of using social media data, we note that the collection of data from participants with verified diagnoses does not come without its own limitations, in addition to its scarcity. Such data can also have biases: clinical texts produced by the clinician and the patient can suffer from reporting bias; observational bias can also be an issue since not all variables involved in predicting a certain clinical outcome are written down in clinical texts; sampling bias can be a serious issue if certain subgroups of a population have limited access to medical or psychological services, etc. There is, therefore, no single best solution and awareness of the limitations of various data sources is crucial. Based on the reviewed literature, such awareness is not always demonstrated and these limitations are not always explicitly discussed.

3.3.2 Applications

While the applications developed within the field are valuable and diverse, a noticeable gap in many of these papers is the lack of discussion on how the research can be translated into public health policies and what risks and ethical implications this process would involve. As NLP technologies for mental disorders demonstrate great potential, we argue that NLP research on this topic should be better aligned to ethical principles guiding the application of AI technology in general (see Mittelstadt (2019)), which includes awareness of how the research will be used in practice and the mitigation of potential risks to ethical principles.

3.3.3 Language Coverage

As noted above, the vast majority of papers returned by our search focused on the English language. This finding, especially since social media data is readily available in various languages, constitutes a serious gap within the field (i.e., see column "language" in Tables 3.1 and 3.2). This gap points to the existence of linguistic inequality in the development of NLP approaches to mental disorders, which could be somewhat rectified (at the very least) by the organisation of shared tasks using non-English social media data. This linguistic coverage gap in the publicly available datasets together with the known difficulties those with a linguistic and/or cultural minority background encounter to access mental healthcare (Kisely 2020; Mianji et al. 2020) make it clear that linguistic inequalities in mental healthcare are a major source of concern.

Another way to address the lack of NLP tools and resources related to mental disorders for non-English languages is to borrow techniques from other NLP domains. For example, a standard approach in NLP when working with a low-resource language is to leverage a machine translation system to translate the task resources available in English into the low-resource language or vice-versa (assuming a publicly available machine translation system for that language pair exists, e.g., Google Translate). This approach is popular in general NLP tasks, e.g., dependency parsing (McDonald et al. 2011) and named entity recognition (Mayhew et al. 2017). In clinical NLP, applications are varied and range from translating search terms for systematic reviews (Balk et al. 2013) to translating EHRs to aid patients who speak foreign languages in the United States (Liu and Cai 2015). For more examples of using machine translation in a clinical NLP setting, we refer the reader to the discussion in Névéol et al. (2018, p. 8).

Recent advances in natural language understanding suggest that a potentially better approach than applying machine translation is possible when dealing with low-resource languages: *cross-lingual transfer learning*. There are many publicly available pretrained multilingual language models (LMs), such as multilingual BERT (Devlin et al. 2019) and XLM-R (Conneau et al. 2020). These multilingual LMs usually cover about a hundred diverse languages (i.e., languages from many language families and with different scripts) and can be adapted relatively easily to different languages and/or tasks. The main idea is to continue training the multilingual LM using task data either in the low-resource language or in English, depending on what is available for the task at hand. Even when task data is available only in English, recent work has shown that pretrained LMs such as XLM-R work well in a *zero-shot cross-lingual* setting on many difficult NLP tasks, i.e., when training the model in one language (e.g., English) and performing inference with the model directly in another language (e.g., the low-resource language). State-of-the-art cross-lingual transfer learning approaches for traditional NLP tasks include performing an add-on training on an intermediate task in English (usually with a lot of data) before fine-tuning the model on the actual task of interest (Phang et al. 2020; Pruksachatkun et al. 2020), and using *adapters*, i.e., small sets of parameters that allow the model to be applied to different languages and tasks (Pfeiffer et al. 2020).

3.3.4 Limitations

One of the main limitations in this work has to do with our choices regarding which papers to include in our survey and which not to. We choose to only use the ACL Anthology as the database where to search for papers, both as a proxy to the "NLP research community" and to keep the number of papers surveyed reasonable. This means we could have missed papers published by NLP researchers, i.e., part of the "NLP community", in venues indexed by other scientific communities. In future work, we plan to extend this search approach to other relevant databases such as the Association Computing Machinery (ACM) or PubMed, which would require the applications of additional filters to ensure that the selected studies use NLP.

We highlight that our intent is not to provide a systematic review. We choose to prioritise breadth of knowledge over depth, and therefore do not aim at discussing specific methods in detail but to provide a snapshot of the current research in NLP for mental health. A detailed look into the specifics of individual disorders is the subject of future work. We also note that the reasons we choose to include some mental disorders and not others are varied: we try to prioritise mental disorders with high burden and epidemiological prevalence (Whiteford et al. 2015), to include sub-areas under the authors' expertise, all while making sure we conform to space constraints. While the issues discussed in this paper reflect the background of the authors (i.e., psychology, NLP, and accessibility), it is possible that some relevant perspectives of professionals from other fields have been omitted or underrepresented.

Overall, the reviewed literature presents an exciting picture of a young and promising multi-disciplinary field. Like any other such field, research on NLP for mental health grapples with problems related to resources, multilingualism, and the placement of its application into a complex real-world context. To solve these issues, the field can benefit from a closer alignment with established data collection and ethical principles in clinical psychology and the broader field of artificial intelligence, as well as a deeper exploration into the use of transfer learning for solving the existing linguistic inequalities within this domain.

3.4 CONCLUSIONS

We provided an overview of the current landscape of the research on the broad theme of NLP for mental health. To do that, we use the ACL Anthology, which is an authoritative database for scientific work in NLP. We survey papers that propose applications and datasets available to address problems related to mental health. The list of datasets provided is flagged as publicly available by their authors, and we note that even if the link to the data is no longer active, as in some cases, the data should in principle still be accessible upon contacting the original authors.

Our focus, in tandem with the core idea in this book, is to highlight solutions available in non-English and low-resource languages. We find that whereas the majority of papers only support English as we initially suspected, possible solutions to address this linguistic inequality include devising shared tasks on mental health-related topics using non-English languages, and more deeply investigating the application of state-of-the-art techniques from NLP such as cross-lingual transfer learning.

NLP technologies create many interesting possibilities to improve mental health services and diagnostics, but there is still a long way to go until this potential can be fully realised. As we seek to achieve a better alignment between the different disciplines that contribute to this type of research, NLP researchers should be mindful of the ethical concerns and linguistic inequalities within the current paradigm of this relatively young research area.

ACKNOWLEDGMENT

Iacer Calixto has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement #838188. Raphael Cardoso has received funding from FAPEG (#202010267000378). We thank our reviewers for their valuable comments.

NOTES

1 https://github.com/iacercalixto/nlp4health
2 For a full list of venues, see https://www.aclweb.org/anthology/
3 https://www.aclweb.org/portal/what-is-cl. Last access on February 1st, 2021.
4 Abbreviation for Autism Spectrum Disorder.
5 Abbreviation for Obsessive Compulsive Disorder.
6 Abbreviation for Attention Deficit Hyperactivity.

REFERENCES

Alamri, M., and W. J. Teahan. 2017. A new error annotation for dyslexic texts in Arabic. In *Proceedings of the Third Arabic Natural Language Processing Workshop*, pages 72–78, Valencia, Spain: Association for Computational Linguistics.

Amblard, M., C. Braud, C. Li, C. Demily, N. Franck, and M. Musiol. 2020. Investigation par méthodes d'apprentissage des spécificités langagières propres aux personnes avec schizophrénie (investigating learning methods applied to language specificity of persons with schizophrenia). In *Actes de la 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Volume 2: Traitement Automatique des Langues Naturelles*, pages 12–26, Nancy, France: ATALA et AFCP.

American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Arlington, VA.

Amir, S., M. Dredze, and J. W. Ayers. 2019. Mental health surveillance over social media with digital cohorts. In *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*, pages 114–120, Minneapolis, Minnesota: Association for Computational Linguistics.

An, M., J. Wang, S. Li, and G. Zhou. 2020. Multimodal topic-enriched auxiliary learning for depression detection. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1078–1089, Barcelona, Spain (Online): International Committee on Computational Linguistics.

Anani, M., N. Kazi, M. Kuntz, and I. Kahanda. 2019. RDoC task at BioNLP-OST 2019. In *Proceedings of The 5th Workshop on BioNLP Open Shared Tasks*, pages 216–226, Hong Kong, China: Association for Computational Linguistics.

Balk, E. M., M. Chung, M. L. Chen, L. K. W. Chang, and T. A. Trikalinos. 2013. Data extraction from machine-translated versus original language randomized trial reports: A comparative study. *Systematic Reviews*, 2, no. 1: 97.

Bar, K., V. Zilberstein, I. Ziv, H. Baram, N. Dershowitz, S. Itzikowitz, and E. Vadim Harel. 2019. Semantic characteristics of schizophrenic speech. In *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*, pages 84–93, Minneapolis, Minnesota: Association for Computational Linguistics.

Barbu, E., M. T. Martín-Valdivia, and L. A. Ureña-López. 2013. Open book: A tool for helping ASD users' semantic comprehension. In *Proceedings of the Workshop on Natural Language Processing for Improving Textual Accessibility*, pages 11–19, Atlanta, Georgia: Association for Computational Linguistics.

Bleiweiss, A. 2020. Neural transduction of letter position dyslexia using an anagram matrix representation. In *Proceedings of the 19th SIGBioMed Workshop on Biomedical Language Processing*, pages 150–155, Online. Association for Computational Linguistics.

Borsboom, D., A. O. Cramer, and A. Kalis. 2019. Brain disorders? not really: Why network structures block reductionism in psychopathology research. *Behavioral and Brain Sciences*, 42: e2.

Boytcheva, S., I. Nikolova, G. Angelova, and Z. Angelov. 2017. Identification of risk factors in clinical texts through association rules. In *Proceedings of the Biomedical NLP Workshop associated with RANLP 2017*, pages 64–72, Varna, Bulgaria: INCOMA Ltd.

Bzdok, D., and A. Meyer-Lindenberg. 2018. Machine learning for precision psychiatry: Opportunities and challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3, no. 3: 223–230.

Canales, L., C. Strapparava, E. Boldrini, and P. Martínez-Barco. 2016. Innovative semi- automatic methodology to annotate emotional corpora. In *Proceedings of the Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media (PEOPLES)*, pages 91–100, Osaka, Japan: The COLING 2016 Organizing Committee.

Chaturvedi, J., N. Viani, J. Sanyal, C. Tytherleigh, I. Hasan, K. Baird, S. Velupillai, R. Stewart, and A. Roberts. 2020. Development of a corpus annotated with medications and their attributes in psychiatric health records. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 2009–2016, Marseille, France: European Language Resources Association.

Cohan, A., B. Desmet, A. Yates, L. Soldaini, S. MacAvaney, and N. Goharian. 2018. SMHD: a large-scale resource for exploring online language usage for multiple mental health conditions. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1485–1497, Santa Fe, New Mexico, USA: Association for Computational Linguistics.

Conneau, A., K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Coppersmith, G., M. Dredze, and C. Harman. 2014. Quantifying mental health signals in Twitter. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 51–60, Baltimore, Maryland, USA: Association for Computational Linguistics.

Delahunty, F., R. Johansson, and M. Arcan. 2019. Passive diagnosis incorporating the PHQ-4 for depression and anxiety. In *Proceedings of the Fourth Social Media Mining for Health Applications (#SMM4H) Workshop & Shared Task*, pages 40–46, Florence, Italy: Association for Computational Linguistics.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota: Association for Computational Linguistics.

Ellendorff, T., S. Foster, and F. Rinaldi. 2016. The PsyMine corpus - a corpus annotated with psychiatric disorders and their etiological factors. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3723–3729, Portorož, Slovenia: European Language Resources Association (ELRA).

Evans, R., C. Orăsan, and I. Dornescu. 2014. An evaluation of syntactic simplification rules for people with autism. In *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)*, pages 131–140, Gothenburg, Sweden: Association for Computational Linguistics.

Fadhil, A., and A. AbuRa'ed. 2019. OlloBot - towards a text-based Arabic health conversational agent: Evaluation and results. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, pages 295–303, Varna, Bulgaria: INCOMA Ltd.

Friðriksdóttir, S. R., and A. K. Ingason. 2020. Disambiguating confusion sets as an aid for dyslexic spelling. In *Proceedings of the 1st Workshop on Tools and Resources to Empower People with REAding DIfficulties (READI)*, pages 1–5, Marseille, France: European Language Resources Association.

Gala, N., and J. Ziegler. 2016. Reducing lexical complexity as a tool to increase text accessibility for children with dyslexia. In *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity (CL4LC)*, pages 59–66, Osaka, Japan: The COLING 2016 Organizing Committee.

Ghosh, S., A. Ekbal, and P. Bhattacharyya. 2020. CEASE, a corpus of emotion annotated suicide notes in English. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 1618–1626, Marseille, France: European Language Resources Association.

Gildea, D., M.-Y. Kan, N. Madnani, C. Teichmann, and M. Villalba. 2018. The ACL Anthology: Current state and future directions. In *Proceedings of Workshop for NLP Open Source Software (NLP-OSS)*, pages 23–28, Melbourne, Australia: Association for Computational Linguistics.

Gkotsis, G., A. Oellrich, T. Hubbard, R. Dobson, M. Liakata, S. Velupillai, and R. Dutta. 2016a. The language of mental health problems in social media. In *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*, pages 63–73, San Diego, CA, USA: Association for Computational Linguistics.

Gkotsis, G., S. Velupillai, A. Oellrich, H. Dean, M. Liakata, and R. Dutta. 2016b. Don't let notes be misunderstood: A negation detection method for assessing risk of suicide in mental health records. In *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*, pages 95–105, San Diego, CA, USA: Association for Computational Linguistics.

Gorrell, G., A. Roberts, R. Jackson, and R. Stewart. 2013. Finding negative symptoms of schizophrenia in patient records. In *Proceedings of the Workshop on NLP for Medicine and Biology associated with RANLP 2013*, pages 9–17, Hissar, Bulgaria: INCOMA Ltd.

Graham, S., C. Depp, E. E. Lee, C. Nebeker, X. Tu, H.-C. Kim, and D. V. Jeste. 2019. Artificial intelligence for mental health and mental illnesses: An overview. *Current Psychiatry Reports*, 21, no. 11: 116.

Gratch, J., R. Artstein, G. Lucas, G. Stratou, S. Scherer, A. Nazarian, R. Wood, J. Boberg, D. DeVault, S. Marsella, D. Traum, S. Rizzo, and L.-P. Morency. 2014. The distress analysis interview corpus of human and computer interviews. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 3123–3128, Reykjavik, Iceland: European Language Resources Association (ELRA).

Gutiérrez, E. D., G. Cecchi, C. Corcoran, and P. Corlett. 2017. Using automated metaphor identification to aid in detection and prediction of first-episode schizophrenia. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2923–2930, Copenhagen, Denmark: Association for Computational Linguistics.

Hauser, M., E. Sariyanidi, B. Tunc, C. Zampella, E. Brodkin, R. Schultz, and J. Parish-Morris. 2019. Using natural conversations to classify autism with limited data: Age matters. In *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*, pages 45–54, Minneapolis, Minnesota: Association for Computational Linguistics.

Heeman, P., R. Lunsford, E. Selfridge, L. Black, and J. van Santen. 2010. Autism and interactional aspects of dialogue. In *Proceedings of the SIGDIAL 2010 Conference*, pages 249–252, Tokyo, Japan: Association for Computational Linguistics.

Holmqvist, M. 2013. *Psychological Disorder*, pages 1554–1556. New York: Springer New York.

Howes, C., M. Purver, R. McCabe, P. G. T. Healey, and M. Lavelle. 2012. Predicting adherence to treatment for schizophrenia from dialogue transcripts. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 79–83, Seoul, South Korea: Association for Computational Linguistics.

Husseini Orabi, A., P. Buddhitha, M. Husseini Orabi, and D. Inkpen. 2018. Deep learning for depression detection of Twitter users. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, pages 88–97, New Orleans, LA: Association for Computational Linguistics.

Ireland, M., and M. Iserman. 2018. Within and between-person differences in language used across anxiety support and neutral Reddit communities. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, pages 182–193, New Orleans, LA: Association for Computational Linguistics.

Jamil, Z., D. Inkpen, P. Buddhitha, and K. White. 2017. Monitoring tweets for depression to detect at-risk users. In *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology — From Linguistic Signal to Clinical Reality*, pages 32–40, Vancouver, BC: Association for Computational Linguistics.

Ji, Y., H. Hong, R. Arriaga, A. Rozga, G. Abowd, and J. Eisenstein. 2014. Mining themes and interests in the asperger's and autism community. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 97–106, Baltimore, Maryland, USA: Association for Computational Linguistics.

Jurafsky, D., and C. Manning. 2012. Natural language processing. *Instructor*, 212, no. 998: 3482.

Just, S., E. Haegert, N. Kořánová, A.-L. Bröcker, I. Nenchev, J. Funcke, C. Montag, and M. Stede. 2019. Coherence models in schizophrenia. In *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*, pages 126–136, Minneapolis, Minnesota: Association for Computational Linguistics.

Kisely, S. 2020. Unravelling the complexities of inequalities in mental healthcare and outcomes for cultural and linguistic minorities. *BJPsych International*, 17, no. 2: 34–37.

Lala, D., P. Milhorat, K. Inoue, M. Ishida, K. Takanashi, and T. Kawahara. 2017. Attentive listening system with backchanneling, response generation and flexible turn- taking. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, pages 127–136, Saarbrücken, Germany: Association for Computational Linguistics.

Lamers, S. M., K. P. Truong, B. Steunenberg, F. de Jong, and G. J. Westerhof. 2014. Applying prosodic speech features in mental health care: An exploratory study in a life- review intervention for depression. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 61–68, Baltimore, Maryland, USA: Association for Computational Linguistics.

Lee, J., T. Cai, W. Xie, and L. Xing. 2020. A counselling corpus in Cantonese. In *Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)*, pages 358–361, Marseille, France: European Language Resources association.

Liu, W., and S. Cai. 2015. Translating electronic health record notes from English to Spanish: A preliminary study. In *Proceedings of BioNLP 15*, pages 134–140, Beijing, China: Association for Computational Linguistics.

Loveys, K., P. Crutchley, E. Wyatt, and G. Coppersmith. 2017. Small but mighty: Affective micropatterns for quantifying mental health from social media language. In *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology — From Linguistic Signal to Clinical Reality*, pages 85–95, Vancouver, BC: Association for Computational Linguistics.

Loveys, K., J. Torrez, A. Fine, G. Moriarty, and G. Coppersmith. 2018. Cross-cultural differences in language markers of depression online. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, pages 78–87, New Orleans, LA: Association for Computational Linguistics.

Lynn, V., A. Goodman, K. Niederhoffer, K. Loveys, P. Resnik, and H. A. Schwartz. 2018. CLPsych 2018 shared task: Predicting current and future psychological health from childhood essays. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, pages 37–46, New Orleans, LA: Association for Computational Linguistics.

Malin, B. A., K. E. Emam, and C. M. O'Keefe. 2013. Biomedical data privacy: Problems, perspectives, and recent advances. Journal of the American Medical Informatics Association. Volume 20, Issue 1, Pages 2–6.

Matero, M., and H. A. Schwartz. 2020. Autoregressive affective language forecasting: A self-supervised task. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2913–2923, Barcelona, Spain (Online): International Committee on Computational Linguistics.

Mayhew, S., C.-T. Tsai, and D. Roth. 2017. Cheap translation for cross-lingual named entity recognition. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2536–2545, Copenhagen, Denmark: Association for Computational Linguistics.

McDonald, R., S. Petrov, and K. Hall. 2011. Multi-source transfer of delexicalized dependency parsers. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 62–72, Edinburgh, Scotland, UK: Association for Computational Linguistics.

Mianji, F., J. Tomaro, and L. J. Kirmayer. 2020. Linguistic and cultural barriers to access and utilization of mental health care for farsi-speaking newcomers in quebec. *International Journal of Migration, Health and Social Care*, 16, no. 4: 495–510.

Mikal, J., S. Hurst, and M. Conway. 2017. Investigating patient attitudes towards the use of social media data to augment depression diagnosis and treatment: A qualitative study. In *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology — From Linguistic Signal to Clinical Reality*, pages 41–47, Vancouver, BC: Association for Computational Linguistics.

Milne, D. N., G. Pink, B. Hachey, and R. A. Calvo. 2016. CLPsych 2016 shared task: Triaging content in online peer-support forums. In *Proceedings of the Third Workshop* *on Computational Linguistics and Clinical Psychology*, pages 118–127, San Diego, CA, USA: Association for Computational Linguistics.

Mitchell, M., K. Hollingshead, and G. 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*, pages 11–20, Denver, Colorado. Association for Computational Linguistics.

Mittelstadt, B. 2019. Principles alone cannot guarantee ethical ai. *Nature Machine Intelligence*, 1, no. 11: 501–507.

Mnookin, S. 2016. Out of the shadows: Making mental health a global development priority (English). http://documents.worldbank.org/curated/en/270131468187759113/Out-of-the-shadows-making-mental-health-a-global-development-priority. Accessed February 3, 2021.

Mowery, D. L., A. Park, C. Bryan, and M. Conway. 2016. Towards automatically classifying depressive symptoms from Twitter data for population health. In *Proceedings of the Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media (PEOPLES)*, pages 182–191, Osaka, Japan: The COLING 2016 Organizing Committee.

Névéol, A., H. Dalianis, S. Velupillai, G. Savova, and P. Zweigenbaum. 2018. Clinical natural language processing in languages other than english: opportunities and challenges. *Journal of Biomedical Semantics*, 9, no. 1: 12.

Parish-Morris, J., E. Sariyanidi, C. Zampella, G. K. Bartley, E. Ferguson, A. A. Pallathra, L. Bateman, S. Plate, M. Cola, J. Pandey, E. S. Brodkin, R. T. Schultz, and B. Tunç. 2018. Oral-motor and lexical diversity during naturalistic conversations in adults with autism spectrum disorder. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, pages 147–157, New Orleans, LA: Association for Computational Linguistics.

Park, S., K. Park, J. Ahn, and A. Oh. 2020. Suicidal risk detection for military personnel. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2523–2531, Online. Association for Computational Linguistics.

Pfeiffer, J., A. Rücklé, C. Poth, A. Kamath, I. Vulić, S. Ruder, K. Cho, and I. Gurevych. 2020. AdapterHub: A framework for adapting transformers. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 46–54, Online. Association for Computational Linguistics.

Phang, J., I. Calixto, P. M. Htut, Y. Pruksachatkun, H. Liu, C. Vania, K. Kann, and S. R. Bowman. 2020. English intermediate-task training improves zero-shot cross-lingual transfer too. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 557–575, Suzhou, China: Association for Computational Linguistics.

Pruksachatkun, Y., J. Phang, H. Liu, P. M. Htut, X. Zhang, R. Y. Pang, C. Vania, K. Kann, and S. R. Bowman 2020. Intermediate-task transfer learning with pretrained language models: When and why does it work? In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5231–5247, Online. Association for Computational Linguistics.

Quiniou, S., and B. Daille. 2018. Towards a diagnosis of textual difficulties for children with dyslexia. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan: European Language Resources Association (ELRA).

Radford, K., L. Lavrencic, R. Peters, K. Kiely, B. Hachey, S. Nowson, and W. Radford. 2018. Can adult mental health be predicted by childhood future-self narratives? Insights from the CLPsych 2018 shared task. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, pages 126–135, New Orleans, LA: Association for Computational Linguistics.

Rauschenberger, M., L. Rello, S. Füchsel, and J. Thomaschewski. 2016. A language resource of German errors written by children with dyslexia. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 83–87, Portorož, Slovenia: European Language Resources Association (ELRA).

Regneri, M., D. King, F. Walji, and O. Palikara. 2020. Images and imagination: Automated analysis of priming effects related to autism spectrum disorder and developmental language disorder. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, pages 11–27, Online. Association for Computational Linguistics.

Rello, L., R. Baeza-Yates, and J. Llisterri. 2014a. DysList: An annotated resource of dyslexic errors. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 1289–1296, Reykjavik, Iceland: European Language Resources Association (ELRA).

Rello, L., H. Saggion, and R. Baeza-Yates. 2014b. Keyword highlighting improves comprehension for people with dyslexia. In *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)*, pages 30–37, Gothenburg, Sweden: Association for Computational Linguistics.

Rello, L., H. Saggion, R. Baeza-Yates, and E. Graells. 2012. Graphical schemes may improve readability but not understandability for people with dyslexia. In *Proceedings of the First Workshop on Predicting and Improving Text Readability for target reader populations*, pages25–32, Montréal, Canada: Association for Computational Linguistics.

Resnik, P., W. Armstrong, L. Claudino, T. Nguyen, V.-A. Nguyen, and J. Boyd-Graber. 2015. Beyond LDA: Exploring supervised topic modeling for depression-related language in Twitter. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 99–107, Denver, Colorado: Association for Computational Linguistics.

Santos, W., A. Funabashi, and I. Paraboni. 2020. Searching Brazilian Twitter for signs of mental health issues. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 6111–6117, Marseille, France: European Language Resources Association.

Shing, H.-C., S. Nair, A. Zirikly, M. Friedenberg, Daumé III, H., and P. Resnik. 2018. Expert, crowdsourced, and machine assessment of suicide risk via online postings. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, pages 25–36, New Orleans, LA: Association for Computational Linguistics.

Song, H., J. You, J.-W. Chung, and J. C. Park. 2018. Feature attention network: Interpretable depression detection from social media. In *Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation*, Hong Kong: Association for Computational Linguistics.

Štajner, S., V. Yaneva, R. Mitkov, and S. P. Ponzetto. 2017. Effects of lexical properties on viewing time per word in autistic and neurotypical readers. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 271–281, Copenhagen, Denmark: Association for Computational Linguistics.

Šuster, S., S. Tulkens, and W. Daelemans. 2017. A short review of ethical challenges in clinical natural language processing. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*, pages 80–87, Valencia, Spain: Association for Computational Linguistics.

Tabak, T., and M. Purver. 2020. Temporal mental health dynamics on social media. In *Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020*, Online. Association for Computational Linguistics.

Tanaka, H., S. Sakti, G. Neubig, T. Toda, and S. Nakamura. 2014. Linguistic and acoustic features for automatic identification of autism spectrum disorders in children's narrative. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 88–96, Baltimore, Maryland, USA: Association for Computational Linguistics.

Viani, N., L. Yin, J. Kam, A. Alawi, A. Bittar, R. Dutta, R. Patel, R. Stewart, and S. Velupillai. 2018. Time expressions in mental health records for symptom onset extraction. In *Proceedings of the Ninth International Workshop on Health Text Mining and Information Analysis*, pages 183–192, Brussels, Belgium: Association for Computational Linguistics.

Ward, A., and R. Crowley. 2011. Story assembly in a dyslexia fluency tutor. In *Proceedings of the Sixth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 130–135, Portland, Oregon: Association for Computational Linguistics.

Whiteford, H. A., A. J. Ferrari, L. Degenhardt, V. Feigin, and T. Vos. 2015. The global burden of mental, neurological and substance use disorders: An analysis from the global burden of disease study 2010. *PloS one*, 10, no. 2: e0116820–e0116820.

World Health Organization. 2018. Mental health: Strengthening our response. https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response. Accessed February 12, 2021.

Yadav, S., J. Chauhan, J. P. Sain, K. Thirunarayan, A. Sheth, and J. Schumm. 2020. Identifying depressive symptoms from tweets: Figurative language enabled multitask learning framework. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 696–709, Barcelona, Spain (Online): International Committee on Computational Linguistics.

Yaneva, V. 2016. *Assessing text and web accessibility for people with autism spectrum disorder*. PhD thesis, University of Wolverhampton. Wolverhampton, United Kingdom.

Yaneva, V., C. Orăsan, R. Evans, and O. Rohanian. 2017. Combining multiple corpora for readability assessment for people with cognitive disabilities. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 121–132, Copenhagen, Denmark: Association for Computational Linguistics.

Yaneva, V., C. Orasan, L. A. Ha, and N. Ponomareva. 2019. A survey of the perceived text adaptation needs of adults with autism. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, pages 1356–1363, Varna, Bulgaria: INCOMA Ltd.

Yaneva, V., I. Temnikova, and R. Mitkov. 2016. A corpus of text data and gaze fixations from autistic and non-autistic adults. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 480–487, Portorož, Slovenia: European Language Resources Association (ELRA).

Yates A., A. Cohan, and N. Goharian. 2017. Depression and self-harm risk assessment in online forums. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2968–2978, Copenhagen, Denmark: Association for Computational Linguistics.

Yip, J. W. C. 2018. Communicating social support in online self-help groups for anxiety and depression: A qualitative discourse analysis. In *Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation*, Hong Kong: Association for Computational Linguistics.

Zirikly, A., P. Resnik, Ö. Uzuner, and K. Hollingshead. 2019. CLPsych 2019 shared task: Predicting the degree of suicide risk in Reddit posts. In *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*, pages 24–33, Minneapolis, Minnesota: Association for Computational Linguistics.