The PsyMine Corpus - A Corpus annotated with Psychiatric Disorders and their Etiological Factors

Tilia Renate Ellendorff, Simon Foster, Fabio Rinaldi
University of Zurich
Institute of Computational Linguistics, Binzmühlestrasse 14, 8050 Zürich, Switzerland
Epidemiology, Biostatistics and Prevention Institute (EBPI), Hirschengraben 84, 8001 Zürich, Switzerland
tilia.ellendorff@uzh.ch, fabio.rinaldi@uzh.ch, simon.foster@uzh.ch

Abstract

We present the first version of a corpus annotated for psychiatric disorders and their etiological factors. The paper describes the choice of text, annotated entities and events/relations as well as the annotation scheme and procedure applied. The corpus is featuring a selection of focus psychiatric disorders including depressive disorder, anxiety disorder, obsessive-compulsive disorder, phobic disorders and panic disorder. Etiological factors for these focus disorders are widespread and include genetic, physiological, sociological and environmental factors among others. Etiological events, including annotated evidence text, represent the interactions between their focus disorders and their etiological factors. Additionally to these core events, symptomatic and treatment events have been annotated. The current version of the corpus includes 175 scientific abstracts. All entities and events/relations have been manually annotated by domain experts and scores of inter-annotator agreement are presented. The aim of the corpus is to provide a first gold standard to support the development of biomedical text mining applications for the specific area of mental disorders which belong to the main contributors to the contemporary burden of disease.

Keywords: Information Extraction, Biomedical Text Mining, Corpus Annotation

1. Introduction

Mental disorders are one of the main contributors to the contemporary burden of disease and one of the major groups of disorders that cause disability (Murray and Lopez, 2013). Therefore, it is important for research in the medical domain to advance in this field. One priority is to discover etiological factors (potential causes) of mental disorders. However, ongoing research faces the problem of having important information scattered across numerous resources, which are mostly textual. Furthermore, mental disorders can have a wide range of underlying factors, spanning over different areas, such as genetic factors, sociological factors, chemical factors and many more. Most of this information is encoded in the written text of research articles of the domain. This makes it hard for researchers to gain an overview of previously discovered causes and therefore obtain a complete picture of a mental disorder and its etiological factors. Biomedical text mining can help by automatically extracting relevant information from written text and presenting it in a structured, more easily accessible format. For building a reliable text mining system, annotated corpora are indispensable. Additionally, systems need to be trained on corpora of the same domain as the target domain in order to show a good performance.

With the PsyMine corpus we present a corpus that is primarily meant to be used for evaluating but could also possibly be used for building a biomedical text mining system for extracting mental disorders together with possible etiological factors. The annotation of the PsyMine is part of the PsyMine project[1]. In the following sections we will describe the corpus, the schema used for annotation, as well as tools that are involved in the annotation process.

2. Related Work

In the past there have been several efforts of corpus annotation in the domain of biomedical text-mining. However, there are not many gold standard corpora for which biomedical events and entities have been manually annotated at word-level. A comprehensive overview of 36 gold standard corpora annotated with entities (and partly events) of the biomedical domain is given in (Neves, 2014). Among the largest and best known biomedical gold standard corpora are for example the Genia Event Extraction Corpus (Kim et al., 2008) and the CRAFT corpus (Bada et al., 2012) featuring various types of biomedical entities. For disease annotations, the Craven corpus (Craven and Kumlien, 1999) and the NCBI disease corpus (Doğan and Lu, 2012) have been widely used in the past. However none of these has a focus on psychiatric disorders.

3. Corpus Selection

The current corpus consists of a total of 175 abstracts of research articles which have been chosen based on the selection of mental disorders described in 4.1.. We selected these abstracts randomly but ensured that all the mental disorders relevant to the PsyMine project are covered, as described in Section 4.1.. The abstracts have been downloaded automatically from PubMed[2], which is the largest database for biomedical research articles, maintained by the US National Library of Medicine (NLM). Each abstract that is part of PubMed is indexed using identifiers provided by MeSH[3] (Medical Subject Headings). These identifiers can be used for retrieving abstract about specific topics. Table 1 shows the number of abstracts per disorder currently registered in PubMed to give an understanding about the overall number

[1] http://www.ontogene.org/current-pr/psymine

[2] http://www.ncbi.nlm.nih.gov/pubmed

[3] https://www.nlm.nih.gov/mesh/

of articles available in PubMed. These numbers have been calculated by checking which articles are associated with the respective MeSH identifiers over the whole of Pubmed. It has to be noted that these are only the current numbers, as new abstracts are continuously being added to PubMed. To ensure that the abstracts and papers are about one of the selected mental disorders and therefore relevant to be included in the corpus, we selected them using the MeSH identifier of the selected disorders on their own, as well as in combination with the [majr] tag. This tag (also provided by MeSH) gives information about the major topics of a PubMed abstract.

We pre-processed these articles through sentence splitting and converted them to a format where each line contains exactly one sentence. This step is supposed to facilitate the annotation process as a less dense way of displaying an article provides a better readability for the human annotators. Within the PsyMine corpus, each abstract has an average of 250 words.

4. Overview of Annotated Entities and Events

The selection of concepts and relations, encoded as events, to be annotated was based on existing terminological resources, and an expert analysis of the specific needs of researchers in the domain of mental health (Center for Mental Health, University of Zurich). Among the resources, the following were considered in particular: International Classification of Diseases (ICD10), Systemized Nomeclature of Medicine (SNOMED), Medical Dictionary of Regulatory Activities (MEDRA), MeSH.

4.1. Selection of Entities

The entities selected for annotation are divided into focus entities and side entities. The entities in focus are a selected set of mental disorders and all etiological factors. Side entity categories include all non-focus mental disorders, all other disorders (non-psychiatric), symptoms of mental disorders as well as treatments of mental disorders. Selected focus psychiatric disorders are organized under a simple hierarchy which is shown in Figure 1. This set of focus psychiatric disorders have been selected for their relevance for the PsyMine project which is concerned with the epidemiology of these specific disorders.

Etiological factors can be loosely grouped according to categories which are illustrated in Table 2.

4.2. Selection of Events

For event annotation, the focus is on etiological events, which were annotated with the event type *associated_with*. Apart from etiological events, the following two side event types were annotated: *is_treatment_for* for disorder-treatment events and *other_relation* for disorder events. Even though etiological events could be further specialized (e.g. scientifically proven correlation versus risk factor), in the current version of the corpus a more general level is maintained. One reason for this decision is that in some of these cases even for domain experts it is hard to decide between a simple correlation and a scientifically proven causal connection.

[4]All number as of October 15, 2015

5. Annotation Schema and Guidelines

The annotation schema is specified by a set of annotation guidelines which are provided to the human annotators. The annotation guidelines describe what exactly should be annotated and how annotations should be made. In the following, a short overview of the most important aspects of the annotation guidelines will be given.

5.1. Entity Annotations

The annotators are asked to annotate all entity mentions belonging to the specified entity types: focus psychiatric disorders, other psychiatric disorders, other disorders, etiological factors, symptoms and treatments as described in Section 4.2.. All entities should be annotated in their standard, nominal form, as well as in their adjectival form or as participles. For example for the concept of *depression*, any of the following variants are annotated: *depression*, *depressive*, *depressed*, *depressive disorder*, *depressive illness*. Furthermore, abbreviated terms referring to entities as well as implicit mentions of entities are also annotated and marked with an *abbrev* attribute or an *implicit* attribute, respectively. Implicit entity mentions are such entity mentions which are mostly anaphoric and implicitly refer to an entity type. Their meaning can only be inferred from the context (e.g. in Figure 3 *the patients* is an implicit entity mention referring to a disorder mentioned before, in this example *unipolar depression*).

5.1.1. Disorder Annotations

Additionally to the variants described above, we decided on including further variants to be annotated as entity mentions of psychiatric disorders in correspondence with the domain experts. We decided to include patient groups, as these are of major importance within research in the field of mental health and in many cases represent a disorder as such. For this reason, we introduced a *patient group* attribute, which, together with the annotation of a psychiatric disorder, marks the annotation of a patient group affected by this disorder. An example of a span of text which would be annotated as a *focus psychiatric disorder* marked with a *patient group* attribute. Example of patient group annotations under the concept of *depression* are *depressive patients* or *patients with depressive disorder*. Furthermore, we introduced a *disorder scale* attribute. In the area of mental health, disorder scales are usually used for referring to a disorder and its severity. In this context they are frequently used in place of disorder mentions. An example for a disorder scale for depression is *Hamilton Depression Rating Scale* (Hamilton, 1960). In many cases, the authors of research paper use abbreviations to refer to disorder scales. If the disorder term itself is part of a patient group or of a disorder rating scale, it is additionally annotated as a separate entity mention. For example *Depression* would be annotated as a separate entity mention within *Hamilton Depression Rating Scale*.

Disorder Type	Relevant MeSH Term	Topic	Major Topic
Depression	Depressive Disorder	84,513	64,588
	Dysthymic Disorder	1,029	620
Anxiety	Anxiety Disorder	72,032	56,049
	Obsessive-Compulsive Disorder	11,839	9,280
	Phobic Disorder	9,524	6,745
	Agoraphobia	2,412	1,396
	Panic Disorder	6,133	4,661

Table 1: Number of Abstracts per Disorder over the whole of PubMed[4]

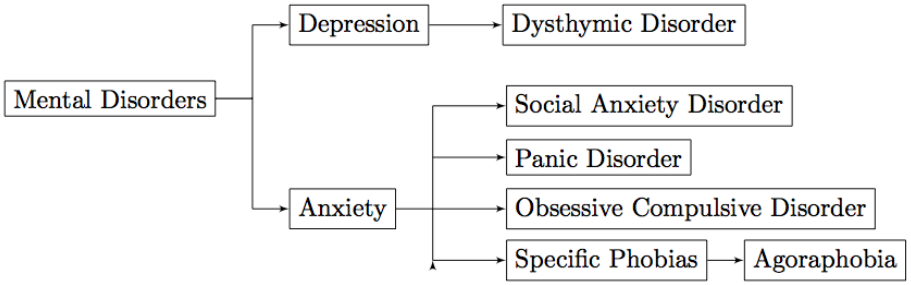

Figure 1: Hierarchy of included Mental Disorders

5.1.2. Etiological Factor Annotations

What to annotate as etiological factors is largely up to the annotator. In research articles of the domain of mental health, the set of possible entity mentions of the etiological factor entity type is an open set in the context of the described annotation task. Nevertheless, all mentions of etiological factors annotated in the corpus can be categorized in any of the categories shown in Table 2. Any span of text annotated as a disorder mention (i.e. focus psychiatric disorder, other psychiatric disorder or other disorder) can at the same time function as an etiological factor and if this is the case, it is additionally annotated as etiological factor. An example for this can be seen in Figure 2.

5.1.3. Treatment and Symptom Annotations

Besides disorders and etiological factors, the PsyMine corpus also includes annotation of treatments and symptoms. The main reason why these entity types are annotated is to be able to differentiate between treatments and symptoms on the one hand and etiological factors on the other hand, as these can show a high level of semantic as well as linguistic similarity.

5.1.4. The Entity Feature Complement

Entity Feature is an additional complimentary entity which is used to add additional important information to entities. One important aspect that is encoded as entity feature are sub-types of disorders, often described through adjectives. Examples are *chronic*, as in *chronic depression* or *treatment-resistant*, as in *treatment-resistant anxiety disorder*. However, in the context of disorder, we decided to exclude a set of variants. These variants, which are always annotated as one item, include multi-word variants which, according to the domain experts, in fact represent one single concept. These include the following: *Major Depression*, *Major Depressive Disorder*, *Panic Disorder with Agoraphobia*, *Panic Disorder without Agoraphobia* and all multi-word items including the word *clinical* (such as *clinical panic*, *clinical anxiety*, etc.).

5.1.5. Anaphora and Equivalence Annotations

As described above, implicit entity mentions are annotated adding the *implicit* attribute to the entity annotation. If an entity marked as implicit has an anaphoric precedent, its connection to the last explicit mention of its anaphoric precedent is marked through an *Anaphora* relation. An example of an implicit entity in combination with an anaphora annotation can be seen in Figure 3. In this example, "the patients" is annotated as Focus Psychiatric Disorder, adding the implicit and patient group attributes. It refers to the previously mentioned "unipolar depressive patients" to which the anaphoric relation is added. If two entity mentions occur in the same abstract which refer to exactly the same concept but which have different surface forms, these two entity mentions are connected through an *Equiv* relation. This is often the case if a text introduces an abbreviation, which then in used in the rest of the article instead of the main variant.

5.2. Relation/Event Annotations

As mentioned above, mentions of all of the following relations types are annotated in the PsyMine corpus: *associated_with* (etiological relations), *treatment_for* (disorder-treatment relations) and *other_relation* (disorder-symptom relations). All relation mentions in the PsyMine corpus are encoded as event, including not only the entities involved but also a span of evidence text. The evidence text is annotated with the type of event and the involved argument entities are connected to the event annotation through the relations of *Cause* and *Theme*. Each event can have several

Etiological Factors	Examples
Genetic_Factors	5HTTLPR
Neurotransmitters	Serotonin
Hormones	Estrogen
Neuronal_Growth_Factors	Brain derived_neurotrophic_factor
Brain_Physiology	Amygdala, Brain_Network, Neuroplasticity
Physiologic_Stress_Response	HPA axis, Cortisol_glucocorticoids
Autonomic_Nervous_System	Anxious Arousal, Autonomic bodily symptoms
Inflammation	Pro inflammatory_Markers, Interleukin_6
Cardiovascular/Vascular_Factors	Heart_rate, Blood_pressure, Hypertension
Disorder_related/Other_Psychiatric_Disorders	Disorder_severity, post traumatic stress disorders
Personality_Factors	Neuroticism
Self_Identity	Core perceptions of the self, self-acceptance associations
Core_Self_Evaluation	Self_esteem
Pregnancy_Birth_factors	Prenatal_maternal_anxiety, Birth_weight
Developmental_Factors	Parenting_style
Cognitive_Factors	Dysfunctional_attitudes
Lifestyle_Factors	Physical_activity, Alcohol_consumption
Sociologic_Factors	Social_status, Income inequality, Discrimination
Socioeconomic_Factors	Education, Income
Socio-demographic_Factors	Age, Gender
Social_Roles	Parenting
Chemicals/Recreational_Drugs	Cigarette_Smoke, Alcohol, Cocaine
Medical_Drugs	Class_SSRI, Escitalopram, Class_SSNRI, Venlafaxine
Environmental_Factors	Aircraft_noise, Amount_of_sunlight

Table 2: Categories of Included Etiological Factors

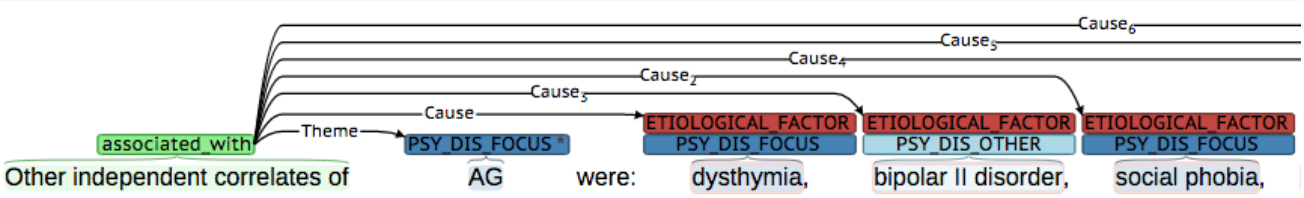

Figure 2: Example of an double annotations etiological factor, psychiatric disorders annotated for (Torres et al., 2014)

Cause and Theme relations that connect to several etiological factors, treatments or symptoms or to several disorders, respectively.

5.2.1. Event Attributes

Event annotations are modified where applicable, using a range of specified event attributes. The attribute *Negation* is used if an event is described as not holding true. The attribute *Speculation* is used if an event is not stated as fact but instead the author merely speculates about an event possibly holding true. The attribute *Revised* is used with events that are described as outdated (i.e. were believed to hold true in the past but have since been revised). Furthermore, the attributes *positive association* and *negative association* are used if the quality of the association is known (e.g. positive correlation versus negative correlation).

6. Annotation Process

The process of corpus annotation for the current version of the corpus was organized in three phases. In the first phase, one domain expert performed a manual annotation, supported by an online annotation tool. The purpose of this annotation step was to clarify and define the annotation scheme and to prove its applicability. The second phase included three annotators and consisted of two sub-steps. The annotators were provided with a documentation of the annotation scheme in the form of annotation guidelines and a two hour training which was followed by a test annotation phase, where each of them annotated 5 abstracts that were carefully selected for relevance to the task. After the test annotation phase, the annotations were checked and the annotators were provided with feedback. In the second substep of the second annotation phase, each annotator annotated 20 abstracts which were then used to measure inter-annotator agreement. In the third phase, the first annotator annotated another 150 abstracts.

The goal of these three phases is to produce a highly reliable reference corpus to be used for validation. In further steps, this corpus will be expanded using a setting of assisted annotation to dramatically improve the speed and effectiveness of the annotation process, as described in section 6.3.. We decided not to apply assisted annotation for the current version of the corpus, so that the annotators are not biased by the system in their decisions and so that the

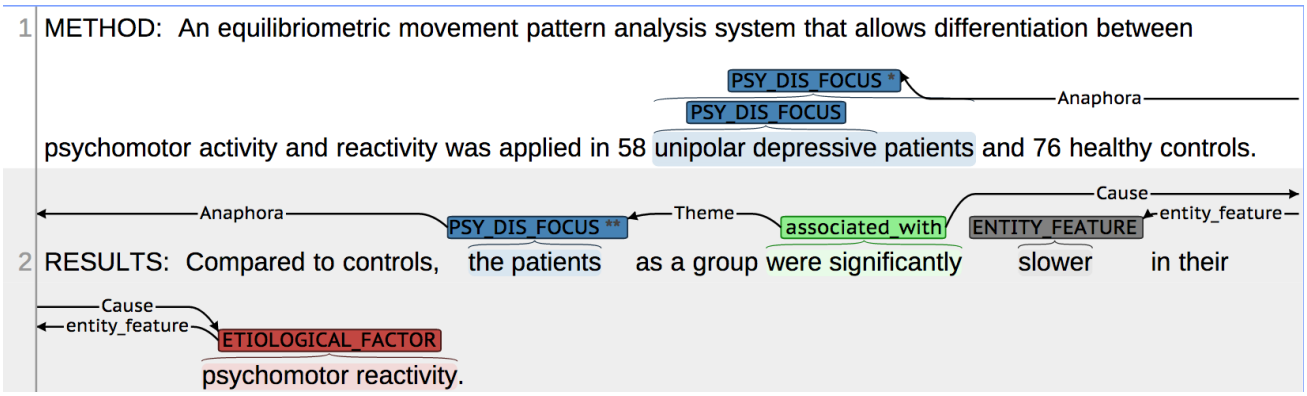

Figure 3: Event Annotation including the annotation of an implicit Entity Mention and an Anaphoric Relation in (Terzivanova and Haralanov, 2014)

Event Type	Event Arguments	Application
associated_with	Cause: Etiological Factor Theme: Psychiatric Disorder (focus + non-focus)	Etiological Events
treatment_for	Cause: Treatment Theme: Psychiatric Disorder (focus + non-focus)	Treatment Events
other_relation	Cause: Symptom/OTHER_ENTITY Theme: Psychiatric Disorder (focus + non-focus)	Symptom Events

Table 3: Overview of annotated Event Types and their Arguments

quality of automatic annotations can be measured against the current version of the corpus as a gold standard.

6.1. Manual Annotation

For manual corpus annotation, we use the brat rapid annotation tool (BRAT)(Stenetorp et al., 2012), which is a web-based annotation tool designed for settings of annotations for natural language processing. BRAT is very intuitive in its application and can be configured for the specific setting of the corpus.

For our purpose, BRAT was configured by defining the entity types described above. Furthermore, we assigned a color scheme to the entities which makes it easier for the annotators to quickly retrieve the entity that they want.

For each relation/event occurrence three separate spans of text are annotated: the two entities, as well as a span of text stating that a relation holds true ("trigger word"). An example of an annotation of this kind can be seen in Figure 4.

We built a converter which can be used to load new PubMed abstracts into BRAT. The current version of the corpus contains purely manual annotations. After completion of manual annotation, all entity mentions of etiological factors were extracted and assigned to the categories shown in Table 2.

6.2. Inter-annotator Agreement

As described above, we calculated inter-annotator agreement (IAA) taking into account 20 abstracts annotated independently by 3 annotators. For focus entities (focus psychiatric disorders and etiological factors), inter-annotator agreement was calculated at the word-level. IAA scores are reported for the focus entity and event type of the PsyMine corpus. For events (etiological events), inter-annotator agreement was calculated at the document-level. All IAA scores were measured in average-observed agreement between annotators and items (AOG). Results can be seen in Table 6. For entities, we applied two different methods of word-level evaluations: strict evaluations and relaxed evaluations. In both cases, only spans of text annotated as entities of the specific entity type were counted. Spans of text that were not annotated by any annotator, are not being considered as matches. Entities that were only annotated by one or two annotators were padded with an auxiliary 'no annotation' category for other annotators. For strict evaluations, the exact spans of text were considered. This means that even entities that were accidentally wrongly annotated with one letter missing or with trailing white-space or punctuation characters included, were counted as a different entity annotation compared to the correctly annotated entity. For the relaxed measure, all overlapping entities where annotators had annotated a slightly different span of text, were counted as matches. Especially for etiological factors, this way of evaluation makes more sense as this set of entities is less defined. For this reason, it is often subject to the own estimation of the annotator, which exact span of text to annotate.

Inter-annotator agreement scores corresponds to the fact that this is not a trivial annotation task and in many cases, not even experts agree in their judgment. Furthermore, as described above, matching non-annotations were not counted as matches which also naturally contributes to a slightly lower score.

Inter-annotator agreement for entity annotations is measured to be higher than inter-annotator agreement for

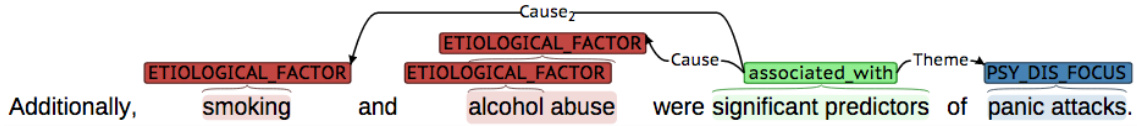

Figure 4: Annotation in Brat of an Etiological Event from (Hearld et al., 2015)

	Annotator 1	Annotator 2	Annotator 3
Focus Psychiatric Disorders	15.5	10.6	15.5
Etiological Factors	22.7	14.3	24.3
Etiological Events	7.8	3.9	4.4

Table 4: Average number of annotated event and entity types per abstract for each annotator

events. One reason for this is that during inter-annotator experiments, one annotator has consistently annotated more events per abstract than the two other annotators, see Table 4. Furthermore, among entity annotations, inter-annotator agreement is much higher for focus disorders compared to etiological factors. The reason of this lies in the nature of these two entity types: whereas focus disorders are a well-defined set, etiological factors can only defined according to the context and in relation to the connected psychiatric disorders.

Furthermore, our analysis consistently shows a notably higher inter-annotator agreement between two of the annotators as compared to the third annotator. This is already visible when comparing average numbers in Table 4. For entities, the values for the relaxed IAA between annotator 1 and 3 reached a score of 0.77 AOG, whereas the IAA between annotators 1 and 2 and annotators 2 and 3 only reached 0.71 and 0.70 AOG, respectively. This difference is even more pronounced for strict evaluation of focus psychiatric disorders, where between annotators 1 and 3 a score of 0.7 AOG is measured, however, between annotators 1 and 2 and annotators 2 and 3 only scores of 0.37 and 0.35 AOG are reached. An overview of IAA between pairs of annotators can be seen in Table 5.

6.3. Assisted Annotation

In an upcoming annotation effort, which is planned to result in a second version of the corpus, we plan to annotate another 200 abstracts automatically (and the remaining full-text papers) before they are checked by the annotators. The advantage is that the annotators only have to accept, reject and complete annotated entities and relations which saves them a lot of work and makes the annotation process faster.

	Entities		**Events**
	Focus Psychiatric Disorders	**Etiological Factors**	**Etiological Events**
strict	0.54	0.43	0.61
relaxed	0.84	0.69	–

Table 6: Inter-annotator agreement (IAA) measured in average observed agreement across coders and items. IAA for entities is measured at the word-level, IAA for events is measured at the document-level

We use a dictionary look-up approach for automatically annotating entities. Our dictionary of mental diseases was compiled from the resources mentioned in section 4.. As there is no vocabulary resource available for etiological factors, the dictionary of etiological factors initially contains the words annotated in step 1. The same holds true for the dictionary of trigger words. Additionally, we extract similar words from the corpus using their vector distance from the words that have been annotated as entities or trigger words and annotate these as well.

7. Corpus Format and Release

The current version is the first version of the PsyMine corpus. An extension is planned in the future. The PsyMine corpus has been made available in two formats. The original BioNLP/BRAT format can be used to display the annotations within the interface of the brat online tool. It is a simple standoff annotation format, first used in the context of the 2009 edition of the BioNLP shared task[5]. Furthermore, it can be used to customize the corpus by making adaptations if necessary for a specific task. The second format is BioC, a standard format for biomedical text mining (Comeau et al., 2013). We also provide a converter between the two formats which is specifically adapted to the entities and relations of the corpus. The corpus will be published online[6].

8. Conclusion

We presented the PsyMine corpus which contains word-level annotations of psychiatric disorders and their etiological factors. The selection of concepts and relations/events included in the corpus is the result of an expert analysis of the specific needs of researchers in the area of mental health.

To our knowledge this is the first corpus compiled specifically for biomedical text mining in the area of psychiatric diseases. The corpus as well as all converter tools involved have been made available to the community. The corpus has been published in BioNLP/BRAT and BioC format as two standard formats for biomedical text mining.

[5] http://www.nactem.ac.uk/tsujii/GENIA/SharedTask/detail.shtml#format

[6] http://www.ontogene.org/current-pr/psymine

		annotator 1 - annotator 2	annotator 2 - annotator 3	annotator 1 - annotator 3
Focus Psychiatric Disorders	strict	0.37	0.35	0.77
	relaxed	0.81	0.81	0.88
Etiological Factors	strict	0.27	0.22	0.38
	relaxed	0.65	0.63	0.69
Both Entity Types	strict	0.27	0.27	0.49
	relaxed	0.71	0.70	0.77

Table 5: Overview of inter-annotator scores between annotator pairs measured in average observed agreement across coders and items

9. Acknowledgments

We would like to thank the three annotators involved in the creation of the PsyMine Corpus for their efforts.
TRE and SF are supported by grant 15-115-R of the COGITO Foundation.

10. Bibliographical References

Bada, M., Eckert, M., Evans, D., Garcia, K., Shipley, K., Sitnikov, D., Baumgartner, W. A., Cohen, K. B., Verspoor, K., Blake, J. A., and Hunter, L. E. (2012). Concept annotation in the craft corpus. *BMC Bioinformatics*, 13(1):1–20.

Comeau, D. C., Dogan, R. I., Ciccarese, P., Cohen, K. B., Krallinger, M., Leitner, F., Lu, Z., Peng, Y., Rinaldi, F., Torii, M., Valencia, A., Verspoor, K., Wiegers, T. C., Wu, C. H., and Wilbur, W. J. (2013). BioC: a minimalist approach to interoperability for biomedical text processing. *Database*, 2013.

Craven, M. and Kumlien, J. (1999). Constructing biological knowledge bases by extracting information from text sources. In *Proceedings of the Seventh International Conference on Intelligent Systems for Molecular Biology*, pages 77–86. AAAI Press.

Doğan, R. I. and Lu, Z. (2012). An improved corpus of disease mentions in pubmed citations. In *Proceedings of the 2012 Workshop on Biomedical Natural Language Processing*, BioNLP '12, pages 91–99, Stroudsburg, PA, USA. Association for Computational Linguistics.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, 1(23):56–62.

Hearld, K. R., Budhwani, H., and Chavez-Yenter, D. (2015). Panic attacks in minority americans: The effects of alcohol abuse, tobacco smoking, and discrimination. *Journal of Affective Disorders*, 174:106 – 112.

Kim, J.-D., Ohta, T., and Tsujii, J. (2008). Corpus annotation for mining biomedical events from literature. *BMC Bioinformatics*, 9(1):1–25.

Murray, C. J. and Lopez, A. D. (2013). Measuring the global burden of disease. *New England Journal of Medicine*, 369(5):448–457. PMID: 23902484.

Neves, M. (2014). An analysis on the entity annotations in biological corpora [version 1; referees: 2 approved]. *F1000Research*, 3(96).

Stenetorp, P., Pyysalo, S., Topić, G., Ohta, T., Ananiadou, S., and Tsujii, J. (2012). Brat: A web-based tool for nlp-assisted text annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, EACL '12, pages 102–107, Stroudsburg, PA, USA. Association for Computational Linguistics.

Terziivanova, P. D. and Haralanov, S. H. (2014). Latent bipolarity in unipolar depression: Experimental findings, conceptual analysis and implications for treatment strategies. *Folia medica*, 56(4):282—288.

Torres, A. R., Ferrão, Y. A., Shavitt, R. G., Diniz, J. B., Costa, D. L., do Rosário, M. C., Miguel, E. C., and Fontenelle, L. F. (2014). Panic disorder and agoraphobia in {OCD} patients: Clinical profile and possible treatment implications. *Comprehensive Psychiatry*, 55(3):588 – 597.