

Using Automated Metaphor Identification to Aid in Detection and Prediction of First-Episode Schizophrenia

E. Darío Gutiérrez¹ Philip R. Corlett² Cheryl M. Corcoran³ Guillermo A. Cecchi¹

¹ IBM Research

² Yale University

³ New York State Psychiatric Institute at Columbia University

edg@icsi.berkeley.edu philip.corlett@yale.edu

cheryl.corcoran@nyspi.columbia.edu guillermo.cecchi@ibm.com

Abstract

The diagnosis of serious mental health conditions such as schizophrenia is based on the judgment of clinicians whose training takes many years and cannot be easily formalized into objective measures. However, clinical research suggests there are disturbances in aspects of the language use of patients with schizophrenia, which opens a door for the use of NLP tools in schizophrenia diagnosis and prognosis. Using metaphor-identification and sentiment-analysis algorithms to automatically generate features, we create a classifier that, with high accuracy, can predict which patients will develop (or currently suffer from) schizophrenia. To our knowledge, this study is the first to demonstrate the utility of automated metaphor identification algorithms for detection or prediction of disease.

1 Introduction

Schizophrenia is a severe mental disorder that has a devastating impact on those who suffer from it, as well as on their families and communities. Schizophrenia is characterized by psychotic behaviors (hallucinations, delusions, thought disorders, movement disorders), flat affect and anhedonia, and trouble with focusing and executive functioning, among other symptoms (American Psychiatric Association, 2013). It afflicts over 21 million people worldwide, and is associated with a 100-150 percent increase in early mortality (Goff et al., 2005; World Health Organization, 2016; Simeone et al., 2015). As a result, diagnosis and treatment of schizophrenia has important public health consequences. Unfortunately, practitioners who are qualified to diagnose and treat serious mental health issues such as schizophrenia are

in chronically short supply, and their accumulated knowledge cannot be easily formalized into reproducible metrics (Patel et al., 2007).

However, clinical research into the symptoms and mechanisms of schizophrenia suggests that disturbances in language use, and especially in metaphor use and affect, characterize schizophrenia. This suggests that automated NLP methods may have the potential to help in diagnosis and prognosis of schizophrenia. In this paper, we work from open-ended transcripts of patients interviewed by non-specialists. We then apply NLP algorithms for metaphor identification and sentiment analysis to automatically generate features for a classifier that, with high accuracy, can predict which patients will develop schizophrenia and which patients would currently be diagnosed with schizophrenia by psychiatrists.

2 Background & Related work

2.1 NLP and Computational Psychiatry

Several recent studies have proven that NLP text-analysis techniques can be successfully applied to predict mental illness. Vincze et al. (2016) use linguistic and demographic features to predict whether a speech transcript was produced by an individual with mild cognitive impairment or by a healthy control. To our knowledge, Elvevåg et al. (2007) were the first to use automated NLP methods to predict whether or not patients suffered from schizophrenia. The technical specifics of their method are unclear, as the paper was intended for a clinical audience, but they use a k -nearest neighbors algorithm in a feature space made up of n -gram features and distributional semantic features to classify 26 schizophrenia patients and 25 healthy controls. They achieve classification accuracy of 78.4% on this task. Mota et al. (2012) employ a graph-based method to classify transcripts taken from interviews with eight patients with

schizophrenia, eight healthy controls, and eight manic patients, achieving both precision and recall of 0.875. Bedi et al. (2015) apply semantic coherence measures and measures based on part-of-speech tags to predict whether 34 youths at risk of psychosis would have a psychotic episode within 2.5 years of being interviewed (five of whom did transition within the study period). They correctly classify 100% of participants.

2.2 Metaphor, Affect and Schizophrenia

Mental-health clinicians have long had intuitions that schizophrenia patients differ from healthy individuals in their use of metaphor. A survey by Kuperberg (2010) of over 50 years of observations in the schizophrenia literature concludes that schizophrenia patients “may use common words in an idiosyncratic or bizarre manner.” Particularly colorful (and metaphorical) examples of bizarre speech recorded by Andreasen (1986) include patients who referred to watches as “time vessels” and to gloves as “hand shoes.”

Billow et al. (1997) carried out the first experimental exploration of this phenomenon. They measure the metaphor production of patients with schizophrenia and healthy controls during free responses to a structured interview. They find that patients with schizophrenia produce comparable rates of felicitous, coherent metaphors as healthy controls, but produce deviant metaphorical speech with significantly greater frequency.

It is not clear what could account for these differences in metaphor production, but neuroscientific studies of patients with schizophrenia offer some clues. Research shows that schizophrenia is associated with dysfunction of the amygdala, a brain structure responsible for regulating emotion (Rasetti et al., 2009). Other work demonstrates impairments in emotion perception and production in patients with schizophrenia (Vaskinn et al., 2008) and even demonstrates that face emotion recognition deficits are a predictor of psychosis onset (Corcoran et al., 2015). Based on these findings, and recognizing the important role that metaphor plays in emotional language (see (Kövecses, 2003)), Elvevåg et al. (2011) hypothesize that metaphor production disturbances in patients with schizophrenia are deeply tied to “emotional” language (i.e., language with high affective polarity). However, it should be noted in this regard that most work on metaphor processing has focused on cortical regions involved (Chen et al., 2008; Schmidt et al., 2010; Benedek et al., 2014).

2.3 Sentiment Analysis & Metaphor Detection Algorithms

Sentiment analysis is a natural-language processing task that involves determining, for given text, whether the text conveys a positive or negative sentiment, and how positive or negative the sentiment is. The book by Liu (2015) gives a comprehensive overview of sentiment analysis.

Metaphor detection is the task of determining whether a given word, phrase, or passage is being used metaphorically or literally. It is an emerging field in NLP, with research still in relatively early stages. A variety of different machine-learning and statistical methods have been applied to the task, including clustering (Birke and Sarkar, 2006; Shutova et al., 2010; Li and Sporleder, 2010; Shutova and Sun, 2013); topic models (Bethard et al., 2009; Li et al., 2010; Heintz et al., 2013); topical structure and imageability analysis (Strzalkowski et al., 2013); semantic similarity graphs (Sporleder and Li, 2009), and feature-based classifiers (Gedigian et al., 2006; Li and Sporleder, 2009; Turney et al., 2011; Dunn, 2013a,b; Hovy et al., 2013; Mohler et al., 2013; Neuman et al., 2013; Tsvetkov et al., 2013, 2014; Klebanov et al.). Metaphor detection methods differ in how they define the task of metaphor detection—for instance, some algorithms seek to determine whether a phrase (such as *sweet victory*) is metaphorical (Krishnakumaran and Zhu, 2007; Turney et al., 2011; Tsvetkov et al., 2014; Bracewell et al., 2014; Gutiérrez et al., 2016), while others attempt to tag metaphoricity at the level of the utterance (Dunn, 2013a), or at the level of individual tokens in running text (Klebanov et al.; Schulder and Hovy, 2014; Do Dinh and Gurevych, 2016). For a recent review, see Shutova (2015). For our purposes, we decided that token-level metaphor detection offered the most appropriate level of granularity, and we chose the algorithm of (Do Dinh and Gurevych, 2016) because of its state-of-the-art performance at this task at the time we began this project.

3 Data

3.1 First-Episode Schizophrenia Transcripts

Our main data set¹ consists of interviews with 17 patients who have suffered a first episode of schizophrenia (denoted by 1EP+) and 15 healthy

¹ Patient data are confidential and can only be used via a Data-Sharing Agreement with authors Corcoran and Corlett; please contact these authors for more information.

controls (denoted by 1EP-). Healthy controls were obtained from the same source population as patients with schizophrenia in the metropolitan New York City region, using web-based advertising on Craigslist, as well as by posting of flyers in and around the region. Participants engaged in open-ended interviews lasting approximately one hour, during which they were encouraged to express themselves narratively. Participants were queried on four topics, for which interviewers provided clarifying questions if they were not spontaneously discussed. The four discussion topics, as well as details of interviewer training and participant selection criteria, are discussed in more detail in the supplementary materials as well as in (Ben-David et al., 2014). Independent transcribers transcribed the interviews. Participants were matched for socioeconomic characteristics and education level. The average age of the 1EP- cohort was 35, and the average age of the 1EP+ cohort was 39. However, the 1EP- cohort was 47% male, while the 1EP+ cohort was 76% male. We refer to this data set as 1EP.

3.2 Prodromal Psychosis Transcripts

We use the data set introduced by Bedi et al. (2015) of transcripts from 34 youths at clinical high risk (CHR) for psychosis, based on the Structured Interview for Prodromal Syndromes (Miller et al., 2003). Demographic details are provided in Bedi et al. (2015). There were no significant differences for age, gender, ethnicity or medication usage between CHR converters vs. CHR non-converters. Notably, all CHR participants were ascertained using gold-standard clinical measures for which the researchers obtained excellent inter-rater reliability with other CHR programs in North America. Open-ended baseline interviews were collected from the participants using the same protocol as above. Participants were then assessed quarterly for 2.5 years to determine whether they had transitioned to psychosis. Five of the participants suffered a first episode of psychosis within the assessment period (denoted by CHR+); the remainder did not (denoted by CHR-).

4 Experiments

The review of the literature in §2.2 suggests that a constellation of disturbances in metaphor use and extremeness/lability of sentiment may characterize schizophrenia. In order to assess whether these phenomena can truly distinguish patients

with schizophrenia from healthy controls or to predict future schizophrenic episodes, we produce five features. Four of these features are derived from sentiment scores produced by a sentiment analysis algorithm, and one is derived from metaphor tags produced by a metaphor identification algorithm.

4.1 Feature Set

Metaphoricity We hope to detect the alteration in metaphor production observed in patients with schizophrenia by Billow et al. (1997) using an automated metaphor detection algorithm that tag word tokens as metaphorical or not. We adapt the token-level metaphor identification algorithm of Do Dinh and Gurevych (2016) to our task. In particular, we use a multilayer perceptron (MLP) architecture with three layers. The input layer is comprised of the concatenation of the word embeddings for each token and the two tokens before and after (not including non-content tokens, and padded with a randomly created embedding at sentence beginnings and endings). The vector for each token is composed of the word’s 300-dimensional Word2Vec skip-gram negative sampling word embedding², concatenated with a one-hot binary vector that indicates the token’s part of speech. The hidden layer has ten fully connected hidden units with the hyperbolic tangent activation function. The output node classifies a token as literal or metaphorical using the softmax activation function.

Training is accomplished by minimizing a cross-entropy objective using stochastic gradient descent; the learning rate is decremented linearly during each epoch, for a maximum of 100 epochs. As in Do Dinh and Gurevych (2016), the MLP is trained on the VU Amsterdam Metaphor Corpus (VUAMC), a subset of the BNC where each token has been annotated as metaphorical or not (Steen et al., 2010), using cross-validation with an 80%-20% train-test split to optimize the regularization and learning rate parameters.

We then measure the percentage of all tokens labeled metaphorical by the metaphor identification algorithm in each transcript, denoting it by Met. We present an example text tagged by this algorithm in figure 1. Notably, the algorithm mistakenly tags the adverbially used preposition *up* in *ended up* as metaphorical; Do Dinh and Gurevych

²<http://drive.google.com/file/d/0B7XkCwpI5KDYNlNUTTlSS21pQmM/>

We ended **up** going to different high schools
...and then at home we also ran in different
social **circles** and things like that.

Figure 1: Sample sentence from one of the transcripts in the 1ep data set. Tokens in **bold** was tagged metaphorical by the token-level metaphor detection algorithm.

(2016) cite this as one of the common failure modes of their algorithm, along with failure to detect metaphors that are only clearly metaphorical from a large amount of surrounding context.

Sentiment We posit that the sentiment scores produced by automated sentiment analysis algorithms should be able to detect disturbances in the production of emotional language, particularly in regard to metaphor. To this end, we create two features that summarize the distribution of sentiment scores in each transcript. In order to obtain token- and phrase-level sentiment scores, we use the implementation of the Recursive Neural Tensor Network sentiment analysis algorithm (Socher et al., 2012) that is included in the Stanford CoreNLP toolkit, with default settings. This implementation comes pre-trained on the Stanford Sentiment Treebank. Tokens are tagged on an integer scale from 1 (Very Negative) to 5 (Very Positive). For each transcript, we take the percentage of all token-level sentiment scores that were either extremely positive (score of 1) or extremely negative (score of 5), which we denote by `SentTok` and similarly compute the percentage of all phrase-level sentiment scores, which we denote by `SentPhr`. We also compute sentiment coherence as

$$\frac{1}{N} \sum_{i=1}^N |s_i - s_{i-1}|$$

where the s_i denotes either the sentiment score for token i (to compute `CohTok`), or the sentiment score for phrase i (to compute `CohPhr`).

4.2 Classification Algorithms

For all algorithms and data sets, we present results produced by leave-one-out cross-validation because of the small number of transcripts available. We use a radial-basis-function support-vector classifier and a convex-hull classifier to classify transcripts based on the variables above. The convex-hull classifier was previously used by Bedi et al. (2015). A test point is classified as originating from a CHR- participant if it lies within

the convex-hull of all the CHR- data points in the training set; otherwise, it is classified as CHR+. The intuition behind the convex-hull approach is that individuals that eventually develop psychosis do not necessarily do so following a unique path to conversion, and moreover psychosis itself cannot be considered a well-defined single condition (Binbay et al., 2012); thus it is reasonable to hypothesize that the “breakdown” of mental abilities may occur along different trajectories for individual CHR+ patients.

5 Results & Discussion

5.1 Statistical Analysis

As predicted, we find that the metaphor identification algorithm does indeed tag a significantly higher proportion of the tokens in the transcripts of patients with schizophrenia as metaphorical (6.3%) than in the healthy controls’ transcripts (5.2%); ($t = 3.76$, $p < .001$). No significant difference was found between the other variables of interest between patients with schizophrenia and healthy controls. No significant difference was found between males and females in metaphor use frequency ($t = 1.105$, $p = 0.28$).

5.2 Classification Performance

First-Episode Schizophrenia Transcripts Table 1 shows the performance of classifiers that individually use each of the five features §4.1 as predictors, as well as the classifier that uses all five in tandem (`All`)³. `Baseline` represents the results of a simple majority classifier (because 18 of the 33 transcripts belonged to patients with schizophrenia, this entails classifying all transcripts as belonging to patients with schizophrenia). Because the 1EP set was not balanced for gender or age, we also present the results of classifying men as having schizophrenia and women as not having schizophrenia (`Gender`) as well as the results of training a classifier on age (`Age`). `Bedi` and `Mota` represent the classification results attained by applying the features/method of Bedi et al. (2015) and Mota et al. (2012), respectively. Using all of the features to train the support-vector classifier performed better than using any of the features individually. The accuracy of the classifier based on all the features was significantly better than baseline (Fisher’s exact test, $p < .005$). Notably, our features outperformed the features

³The `All` classifier does not use gender or age features.

suggested by Bedi et al. (2015) and by Mota et al. (2012).

Prodromal Psychosis Transcripts On the prodromal transcripts, a classifier trained on all the features once again outperformed classifiers on any of the features individually, which performed at or near baseline. Interestingly, the convex-hull classifier outperformed the support vector classifier on this data. The convex-hull classifier trained on all five features correctly identified the outcome of 33 of the 34 CHR patients (97.1% accuracy). The sole patient who was misclassified belonged to the CHR+ group. This is comparable to the 100% accuracy of the Bedi et al. (2015) method and superior to the 79.4% accuracy of the Mota et al. (2012) method.

In order to explore the relationship between the two data sets, we also applied the best classifier trained on the 1EP data to the prodromal data. Interestingly, the 1EP classifier tagged 29 of the 34 CHR patients as patients with schizophrenia, including all five patients in the CHR+ group. The hypersensitivity of the 1EP classifier when applied to the prodromal data suggests that the cues that discriminate between patients with first-episode schizophrenia and healthy controls tend to place CHR patients into the same category as patients with first-episode schizophrenia. It is worth noting that the classifier tagged all of the CHR+ patients as 1EP+. We believe this indicates that our method would be useful as a tool meant to channel limited attention and resources toward patients with particularly high risk (above and beyond the criteria that currently flag a patient as being CHR).

6 Conclusion

To our knowledge, this study is the first to demonstrate the utility of automated metaphor identification algorithms in a public-health setting,

and particularly for the prediction or detection of schizophrenia. Our algorithm’s performance on the task of schizophrenia diagnosis from transcripts outperforms the two existing methods detailed in existing literature.

Our results also contribute to clinical knowledge of the nature of language-use abnormalities in schizophrenia, as they support previous research which finds that those suffering from schizophrenia produce more metaphors in free speech than healthy controls. Previously it was only possible to measure such disturbances by labor-intensive and subjective hand-coding of transcripts for metaphoricality, or by the assessment of expert clinicians, whose time is limited. This work breaks new ground by showing that such disturbances can be measured in an automated and reproducible fashion, using features generated via machine learning.

Our work is somewhat constrained by the small sample size available to us. As our data comes from a vulnerable population, obtaining a larger data set is challenging, but essential for future work. In fact, two of the authors are in the process of collecting data from a total of 120 CHR individuals. This would enable a more thorough investigation of a larger and more sophisticated suite of linguistic features, and especially a more fine-grained analysis of the interaction of metaphor and emotional language in schizophrenia.

Acknowledgments

This work was funded in part via NIMH grants R01MH107558 and R03MH108933. The authors have no conflicts to declare.

References

- American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. American Psychiatric Association.
- Nancy C. Andreasen. 1986. The scale for assessment of thought, language and communication (TLC). *Schizophrenia Bulletin* 12:473–482.
- Gillinder Bedi, Facundo Carrillo, Guillermo A. Cecchi, Diego Fernández Slezak, Mariano Sigman, Natália B. Mota, Sidarta Ribeiro, Daniel C. Javitt, Mauro Copelli, and Cheryl M. Corcoran. 2015. Automated analysis of free speech predicts psychosis onset in high-risk youths. *npj Schizophrenia* 1:15030.
- S. Ben-David, M. Birnbaum, M. Eilenberg, J. DeVlyder, K. Gill, J. Schienle, N. Azimov, E.P. Lukens,

Table 1: Classification performance on 1ep set.

Variables	F-score	Accuracy
Baseline	0.703	0.563
Gender	0.703	0.656
Age	0.629	0.594
CohTok	0.694	0.531
CohPhr	0.703	0.656
Met	0.789	0.750
SentTok	0.732	0.688
SentPhr	0.718	0.656
All	0.848	0.844
Bedi	0.744	0.688
Mota	0.732	0.688

- L. Davidson, and C.N. Corcoran. 2014. The subjective experience of youths at clinical high risk for psychosis: a qualitative study. *Psychiatric Services* pages 1499–1501.
- Mathias Benedek, Roger Beaty, Emanuel Jauk, Karl Koschutnig, Andreas Fink, Paul J. Silvia, Beate Dunst, and Aljoscha C. Neubauer. 2014. The neural basis of figurative language production. *NeuroImage* 90:99–106.
- Steven Bethard, Vicky Tzuyin Lai, and James H. Martin. 2009. Topic model analysis of metaphor frequency for psycholinguistic stimuli. In *Proceedings of the Workshop on Computational Approaches to Linguistic Creativity*. Association for Computational Linguistics, pages 9–16.
- Richard M Billow, Jeffrey Rossman, Nona Lewis, Deborah Goldman, and Charles Raps. 1997. Observing expressive and deviant language in schizophrenia. *Metaphor and Symbol* 12(3):205–216.
- T. Binbay, M. Drukker, H. Elbi, F.A. Tank, F. zknay, H. Onay, N. Zal, J. van Os, and K. Alptekin. 2012. Testing the psychosis continuum: differential impact of genetic and nongenetic risk factors and comorbid psychopathology across the entire spectrum of psychosis. *Schizophrenia Bulletin* 38:992–1002.
- Julia Birke and Anoop Sarkar. 2006. A clustering approach for nearly unsupervised recognition of non-literal language. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*. pages 329–336.
- David B. Bracewell, Marc T. Tomlinson, Michael Mohler, and Bryan Rink. 2014. A tiered approach to the recognition of metaphor. In *International Conference on Intelligent Text Processing and Computational Linguistics (CICLing)*. pages 403–414.
- Evan Chen, Page Widick, and Anjan Chatterjee. 2008. Functional-anatomical organization of predicate metaphor processing. *Brain and Language* 107:194–202.
- C.M. Corcoran, J.G. Keilp, J. Kayser, C. Klim, P.D. Butler, G.E. Bruder, R.C. Gur, and D.C. Javitt. 2015. Emotion recognition deficits as predictors of transition in individuals at clinical high risk for schizophrenia: a neurodevelopmental perspective. *Psychological Medicine* 45:2959–2973.
- Erik-Lân Do Dinh and Iryna Gurevych. 2016. Token-level metaphor detection using neural networks. In *Proceedings of the Fourth Workshop on Metaphor in NLP*. Association for Computational Linguistics, San Diego, California, pages 28–33.
- Jonathan Dunn. 2013a. Evaluating the premises and results of four metaphor identification systems. In *Computational Linguistics and Intelligent Text Processing*, Springer, pages 471–486.
- Jonathan Dunn. 2013b. What metaphor identification systems can tell us about metaphor-in-language. In *Proceedings of the First Workshop on Metaphor in NLP*. pages 1–10.
- Brita Elvevåg, Peter W. Foltz, Daniel R. Weinberger, and Terry E. Goldberg. 2007. Quantifying incoherence in speech: An automated methodology and novel application to schizophrenia. *Schizophrenia Research* 93(1):304–316.
- Brita Elvevåg, Kim Helsen, Marc De Hert, Kim Sweers, and Gert Storms. 2011. Metaphor interpretation and use: a window into semantics in schizophrenia. *Schizophrenia Research* 133(1):205–211.
- Matt Gedigian, John Bryant, Srini Narayanan, and Branimir Cicic. 2006. Catching metaphors. In *Proceedings of the Third Workshop on Scalable Natural Language Understanding*. Association for Computational Linguistics, New York, pages 41–48.
- D. C. Goff, C. Cather, A.E. Evins, D.C. Henderson, O. Freudenreich, P.M. Copeland, and F.M. Sacks. 2005. Medical morbidity and mortality in schizophrenia: guidelines for psychiatrists. *Journal of Clinical Psychiatry* 66:183–194.
- E. Darío Gutiérrez, Ekaterina Shutova, Tyler Marghetis, and Benjamin K. Bergen. 2016. Literal and metaphorical senses in compositional distributional semantic models. In *Annual Meeting of the Association for Computational Linguistics*. pages 183–193.
- Ilana Heintz, Ryan Gabbard, Mahesh Srinivasan, David Barner, Donald S Black, Marjorie Freedman, and Ralph Weischedel. 2013. Automatic extraction of linguistic metaphor with lda topic modeling. In *Proceedings of the First Workshop on Metaphor in NLP*. pages 58–66.
- Dirk Hovy, Shashank Srivastava, Sujay Kumar Jauhar, Mrinmaya Sachan, Kartik Goyal, Huiying Li, Whitney Sanders, and Eduard Hovy. 2013. Identifying metaphorical word use with tree kernels. In *Proceedings of the First Workshop on Metaphor in NLP*. pages 52–57.
- Beata Beigman Klebanov, Chee Wee Leong, Michael Heilman, and Michael Flor. 2007. Different texts, same metaphors: unigrams and beyond. In *Proceedings of the Second Workshop on Metaphor in NLP*. Baltimore, MD, USA, pages 11–17.
- Zoltán Kövecses. 2003. *Metaphor and emotion: Language, culture, and body in human feeling*. Cambridge University Press, Cambridge, UK.
- Saisuresh Krishnakumaran and Xiaojin Zhu. 2007. Hunting elusive metaphors using lexical resources. In *Proceedings of the Workshop on Computational approaches to Figurative Language*. Association for Computational Linguistics, pages 13–20.

- Gina R. Kuperberg. 2010. Language in schizophrenia part 1: an introduction. *Language and Linguistics Compass* 4(8):576–589.
- Linlin Li, Benjamin Roth, and Caroline Sporleder. 2010. Topic models for word sense disambiguation and token-based idiom detection. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, pages 1138–1147.
- Linlin Li and Caroline Sporleder. 2009. Classifier combination for contextual idiom detection without labelled data. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*. Association for Computational Linguistics, pages 315–323.
- Linlin Li and Caroline Sporleder. 2010. Using Gaussian mixture models to detect figurative language in context. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, pages 297–300.
- Bing Liu. 2015. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge University Press, Cambridge, UK.
- Tandy J. Miller, Thomas H. McGlashan, Joanna L. Rosen, Kristen Cadenhead, Joseph Ventura, William McFarlane, Diana O. Perkins, Godfrey D. Pearson, and Scott W. Woods. 2003. Prodromal assessment with the structured interview for prodromal syndromes and the scale of prodromal symptoms: predictive validity, interrater reliability, and training to reliability. *Schizophrenia Bulletin* 29:703–715.
- Michael Mohler, David Bracewell, David Hinote, and Marc Tomlinson. 2013. Semantic signatures for example-based linguistic metaphor detection. In *Proceedings of the First Workshop on Metaphor in NLP*. pages 27–35.
- Natalia B Mota, Nivaldo AP Vasconcelos, Nathalia Lemos, Ana C Pieretti, Osame Kinouchi, Guillermo A Cecchi, Mauro Copelli, and Sidarta Ribeiro. 2012. Speech graphs provide a quantitative measure of thought disorder in psychosis. *PLoS one* 7(4):e34928.
- Yair Neuman, Dan Assaf, Yohai Cohen, Mark Last, Shlomo Argamon, Newton Howard, and Ophir Frieder. 2013. Metaphor identification in large texts corpora. *PLoS ONE* 8:e62343.
- Vikram Patel, Alan J Flisher, Sarah Hetrick, and Patrick McGorry. 2007. Mental health of young people: a global public-health challenge. *The Lancet* 369(9569):1302–1313.
- Roberta Rasetti, Venkata S. Mattay, Lisa M. Wiedholz, Bhaskar S. Kolachana, Ahmad R. Hariri, Joseph H. Callicott, Andreas Meyer-Lindenberg, and Daniel R. Weinberger. 2009. Evidence that altered amygdala activity in schizophrenia is related to clinical state and not genetic risk. *American Journal of Psychiatry* 166(2):216–225.
- Gwenda L. Schmidt, Alexander Kranjec, Eileen R. Cardillo, and Anjan Chatterjee. 2010. Beyond laterality: a critical assessment of research on the neural basis of metaphor. *Journal of the International Neuropsychological Society* 16:1–5.
- Marc Schulder and Eduard Hovy. 2014. Metaphor detection through term relevance. In *Proceedings of the Second Workshop on Metaphor in NLP*. Baltimore, MD, USA, pages 18–26.
- Ekaterina Shutova and Lin Sun. 2013. Unsupervised metaphor identification using hierarchical graph factorization clustering. In *Human Language Technologies: The 2013 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. pages 978–988.
- Ekaterina Shutova, Lin Sun, and Anna Korhonen. 2010. Metaphor identification using verb and noun clustering. In *Proceedings of the 23rd International Conference on Computational Linguistics*. Association for Computational Linguistics, pages 1002–1010.
- Ekaterina Shutova. 2015. Design and evaluation of metaphor processing systems. volume Forthcoming.
- Jason C Simeone, Alexandra J Ward, Philip Rotella, Jenna Collins, and Ricarda Windisch. 2015. An evaluation of variation in published estimates of schizophrenia prevalence from 1990–2013: a systematic literature review. *BMC psychiatry* 15(1):193.
- Richard Socher, Brody Huval, Christopher D. Manning, and Andrew Y. Ng. 2012. Semantic compositionality through recursive matrix-vector spaces. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. Association for Computational Linguistics, pages 1201–1211.
- Caroline Sporleder and Linlin Li. 2009. Unsupervised recognition of literal and non-literal use of idiomatic expressions. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, pages 754–762.
- Gerard J. Steen, Aletta G. Dorst, J. Berenike Herrmann, Anna Kaal, Tina Krennmayr, and Trijntje Pasma. 2010. *A method for linguistic metaphor identification: From MIP to MIPVU*, volume 14. John Benjamins Publishing, Amsterdam/Philadelphia.
- Tomek Strzalkowski, George A. Broadwell, Sarah Taylor, Laurie Feldman, Boris Yamrom, Samira Shaikh, Ting Liu, Kit Cho, Umit Boz, Ignacio Cases, and Kyle Elliot. 2013. Robust extraction of metaphors

from novel data. In *Proceedings of the First Workshop on Metaphor in NLP*. Association for Computational Linguistics, Atlanta, Georgia, pages 67–76.

Yulia Tsvetkov, Leonid Boytsov, Anatole Gershman, Eric Nyberg, and Chris Dyer. 2014. Metaphor detection with cross-lingual model transfer. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. pages 248–258.

Yulia Tsvetkov, Elena Mukomel, and Anatole Gershman. 2013. Cross-lingual metaphor detection using common semantic features.

Peter D. Turney, Yair Neuman, Dan Assaf, and Yohai Cohen. 2011. Literal and metaphorical sense identification through concrete and abstract context. In *Proceedings of the 2011 Conference on the Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Stroudsburg, PA, USA, EMNLP ’11, pages 680–690.

Anja Vaskinn, Kjetil Sundet, Svein Friis, Carmen Simonsen, Astrid B Birkenaes, Halldora JONsdOTTir, Petter Andreas Ringen, and Ole A Andreassen. 2008. Emotion perception and learning potential: mediators between neurocognition and social problem-solving in schizophrenia? *Journal of the International Neuropsychological Society* 14(02):279–288.

Veronika Vincze, Gábor Gosztolya, László Tóth, Ildikó Hoffmann, and Gréta Szatlóczki. 2016. Detecting mild cognitive impairment by exploiting linguistic information from transcripts. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Berlin, Germany, pages 181–187.

World Health Organization.
2016. [Schizophrenia fact sheet](http://www.who.int/mediacentre/factsheets/fs397/en/).
<http://www.who.int/mediacentre/factsheets/fs397/en/>.