Psychiatry Research 225 (2015) 40–49

Contents lists available at ScienceDirect

Psychiatry Research

journal homepage: www.elsevier.com/locate/psychres

Lexical use in emotional autobiographical narratives of persons with schizophrenia and healthy controls

Kai Hong [a], Ani Nenkova [a], Mary E. March [b], Amber P. Parker [c], Ragini Verma [d], Christian G. Kohler [b,*]

[a] Department of Computer and Information Science, University of Pennsylvania School of Engineering and Applied Science, USA
[b] Schizophrenia Research Center, Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA
[c] University of Pennsylvania School of Art and Sciences, Philadelphia, PA 19104, USA
[d] Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104, USA

ARTICLE INFO

Article history:
Received 7 October 2013
Received in revised form
26 September 2014
Accepted 4 October 2014
Available online 3 December 2014

Keywords:
Emotion
Lexical features
LIWC
Diction
Learning-based analyses
Text classification

ABSTRACT

Language dysfunction has long been described in schizophrenia and most studies have focused on characteristics of structure and form. This project focuses on the content of language based on autobiographical narratives of five basic emotions. In persons with schizophrenia and healthy controls, we employed a comprehensive automated analysis of lexical use and we identified specific words and semantically or functionally related words derived from dictionaries that occurred significantly more often in narratives of either group. Patients employed a similar number of words but differed in lower expressivity and complexity, more self-reference and more repetitions. We developed a classification method for predicting subject status and tested its accuracy in a leave-one-subject-out evaluation procedure. We identified a set of 18 features that achieved 65.7% accuracy in predicting clinical status based on single emotion narratives, and 74.4% accuracy based on all five narratives. Subject clinical status could be determined automatically more accurately based on narratives related to anger or happiness experiences and there were a larger number of lexical differences between the two groups for these emotions compared to other emotions.

© 2014 Published by Elsevier Ireland Ltd.

1. Introduction

Narratives of emotional experiences contain rich personal information in linguistic form. It has been suggested that evolutionary brain development has led to structural asymmetries that underlie the components of human language (Geschwind and Galaburda, 1987), and that these asymmetries relate to the emergence of schizophrenia as a human disorder (Crow, 1997). Disturbed functions of language have long been described in persons with schizophrenia, for example, in areas of phonetic intonation (dysprosody), lack of volume or content (alogia) and disturbed content or relatedness of speech production (thought disorder or schizophasia). Considered as core phenomenologic characteristics of the illness, these functions relate to different linguistic categories, as summarized in reviews (DeLisi, 2001; Covington et al., 2005; McKenna and Oh, 2005). Examinations of spontaneous and conversational speech in schizophrenia have tended to focus on measures of coherence, which represents the semantic relationship of expressed ideas. Later analyses include cloze procedure (Manschreck et al., 1981; Newby, 1998), ambiguous referents (Docherty et al., 1996) and unusual word combinations (Solovay et al., 1987; Niznikiewicz et al., 2002). All these studies have detected abnormalities related to the production of coherent discourse. The findings hold even when the comparison group is patients with mood disorders rather than healthy persons (Docherty et al., 1996). These prior studies have also confirmed the relationship between peculiarities of language use and cognitive dysfunction involving attention and executive abilities (Docherty, 2005; Marini et al., 2008).

Despite these compelling findings, analysis of patent's language production on a large scale remains a practical problem. Human annotation of patient speech is time consuming and rather subjective for some categories of linguistic expression. Patients often do not talk much, so there is relatively short sample of language available for the analysis. To address these challenges in prior work, Elvevåg et al. (2007) applied *Latent Semantic Analysis* (*LSA*) to compute automatically semantic similarities between the answers to a set of standardized questions involving different

* Correspondence to: Neuropsychiatry Section, Department of Psychiatry, 10th Floor, Gates Building, University of Pennsylvania School of Medicine, 3400 Spruce Street, Philadelphia, PA 19104, USA. Tel.: +1 215 614 0161; fax: +1 215 662 7903.
E-mail addresses: hongkai1@seas.upenn.edu (K. Hong), nenkova@seas.upenn.edu (A. Nenkova), memarch@mail.med.upenn.edu (M.E. March), parker@sas.upenn.edu (A.P. Parker), ragini.verma@gmail.com (R. Verma), kohler@mail.med.upenn.edu (C.G. Kohler).

http://dx.doi.org/10.1016/j.psychres.2014.10.002
0165-1781/© 2014 Published by Elsevier Ireland Ltd.

thematic areas. The similarities were computed at three levels of granularity: words, sentences and entire answers. In that study, *LSA* scores for answer coherence correlated reasonably well with human ratings of thought disorder. The proposed method was able to discriminate between patients with high levels of thought disorder from controls. Cretchley et al. (2010) conducted qualitative analysis by examining the conversations between caretakers and schizophrenic patients using Leximancer. The Leximancer toolkit employs word-association information to extract concepts. This method makes it feasible to generate a tailored taxonomy for each dataset (Smith, 2003). Cretchley et al. (2010) found that the carers used different strategies to communicate with the schizophrenia patients, depending on the conversational tendencies and relationship context of the patients. In our work we focus exclusively on examining lexical features in personal narratives about past experiences related to five basic emotions. Our goal is to identify differences in lexical use between schizophrenia patients and healthy controls.

We explore four types of lexical features to characterize narratives: *generic*, *word identity*, *dictionary* and *language model* (LM) features. *Generic* features include the number of words per sentence, the number of letters (graphemes) per word, the number of sentences per narrative and the number of repetitions of any words. Traditionally *generic* features have been used in the readability literature, in the context of a task to find reading material appropriate for a given grade level of reading competency (Heilman et al., 2007). They reflect the complexity of the analyzed language. *Word identity* features, on the other hand, track the frequency of individual words. These features have been employed in analyses of cancer concerns (Ando et al., 2007) and detection of post-traumatic stress disorder (He et al., 2012). In work specifically related to schizophrenia, word identity features derived from outpatient consultations between patients and psychiatrists have been shown useful for predicting positive and negative syndrome scale (PANSS) and adherence to treatment for schizophrenic patients (Howes et al., 2012a, 2012b, 2013).

Dictionary features provide a more robust way to analyze lexical use by grouping words into semantic and grammatical categories. One can track the occurrences of these more interpretable and general categories rather than the occurrences of individual words. A popular dictionary-based package for psychometric analysis is the *Linguistic Inquiry and Word Count* (*LIWC*) (Pennebaker et al., 2007), which groups words into psychologically meaningful categories. *LIWC* has been applied to analyses of various forms of texts, including written accounts of personal emotional experiences and transcripts of spoken narratives. It has been shown that the use of pronouns and function words are good indicators to reveal if the narrator is deceptive or honest (Newman et al., 2003), what is the narrator's personality type (Pennebaker and King, 1999) and mental health status (Tausczik and Pennebaker, 2010). *LIWC* and *word identity* features can elucidate personality styles, such as *introversion*, *openness*, and *conscientiousness* from analysis of e-mails (Gill et al., 2006), blogs (Gill et al., 2009) and conversations (Mairesse et al., 2007). One of the main findings based on LIWC dictionary features is that the increased use of first-person pronouns is a reliable indicator of emotional distress (Rude et al., 2004) and suicide intent (Stirman and Pennebaker, 2001).

Finally we experiment with *language model* (LM) features. In these models the probability of word sequences (*n*-grams) is first estimated from a training corpus. We train one LM for patients and one for controls, using the narratives from the corresponding groups as training data respectively. After that, the LMs are used to estimate the likelihood of new narratives, effectively summarizing in a single number the similarity of the narrative to the previously seen narratives from each of the two groups. In previous research LMs have been used to detect language dominance in bilingual children (Solorio et al., 2011), autism (Prud'hommeaux et al., 2011) and language impairment in monolingual and bilingual children (Gabani et al., 2009).

Our project focuses on discovering differences in lexical use between persons with schizophrenia and healthy controls, based on narratives in which the subjects described past emotional experiences that evoked happy, sad, angry, fear and disgust emotions. We performed automated analyses of *generic*, *word identity*, *dictionary* and *LM* features in the autobiographical narratives of the subjects in our study. From the full set of available features we identified a limited set of features that were sufficient to distinguish the clinical status of the subject. We observed that persons with schizophrenia offer autobiographical descriptions of experiences that consist of: (1) shorter words, fewer words per sentence and more sentences per narrative, (2) a greater number of references to themselves, (3) a higher number of word repetition and (4) adverbs that denote intensity which were included in the psychiatrist's questions. We developed a comprehensive machine learning model by identifying a set of lexical features for each training fold. Our supervised classifier can differentiate persons with schizophrenia from normal controls with 74.4% accuracy. To examine the effectiveness of each feature group, we performed ablation experiments. It turned out that removing each feature group led to a decrease in performance, with the largest decrease occurring when word identity features were removed. The significant features detected in our work provide a candidate list of linguistic features that can be tested robustly in future work.

2. Method

2.1. Participants

As part of our standardized method of obtaining evoked facial expressions (Kohler et al., 2010), we collected autobiographical narratives from 39 participants (19 male, 20 female; 23 persons with schizophrenia and 16 controls), group matched for age, gender and ethnicity. Demographic and clinical information is provided in Table 1. All patients were deemed clinically stable, without hospitalization for the past 6 months or no change in antipsychotic medications for the past 3 months.

Participants were asked to narrate their autobiographical experiences of five universal emotions: anger, disgust, fear, happy and sad in pseudorandomized order. Subjects were instructed "to narrate the experience, as if telling to a good friend", describing the general setting and proceeding to include moments when they experienced the target emotion to a mild, moderate or extreme degree. Narratives lasted between 30 and 90 s and contained 188 words on average. We obtained 120 narratives from patients and 81 from controls, as several participants offered

Table 1
Demographic information for schizophrenic patients and healthy controls

Variables	Full sample ($N=39$)	Schizophrenia ($N=23$)	Control ($N=16$)
Gender: male	19	12	7
Ethnic			
White	12	8	4
Black/African American	23	12	11
Asian/Hybrid	4	3	1
Mean age (S.D.)	33.21 (8.58)	33.81 (9.65)	32.29 (6.59)
Education (S.D.)	13.63 (2.21)	13.08 (2.21)	14.47 (2.00)
Mother education (S.D.)	13.39 (3.12)	13.33 (3.40)	13.06 (2.79)
Father education (S.D.)	13.63 (3.33)	13.65 (3.83)	13.71 (2.58)
SANS (S.D.)***	18.90 (17.22)	28.04 (15.85)	4.87 (6.40)
SAPS (S.D.)***	7.82 (12.03)	12.52 (13.40)	0.6 (2.24)
HAM-D (S.D.)		5.95 (4.78)	
LevFun (S.D.)***	26.75 (8.31)	24.52 (8.07)	33.43 (4.96)

None of the other differences were significant at the 95% confidence level.

*** $p < 0.001$ was noted for two-tailed t-test p-values.

Download English Version:

https://daneshyari.com/en/article/332078

Download Persian Version:

https://daneshyari.com/article/332078

Daneshyari.com