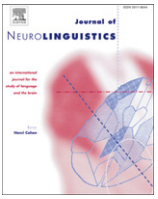
Journal of Neurolinguistics 23 (2010) 270–284

Contents lists available at ScienceDirect

Journal of Neurolinguistics

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



An automated method to analyze language use in patients with schizophrenia and their first-degree relatives

Brita Elvevåg [a,*], Peter W. Foltz [b,c], Mark Rosenstein [b], Lynn E. DeLisi [d,e]

[a] Clinical Brain Disorders Branch, National Institute of Mental Health, National Institutes of Health, Building 10, Room 3C104, MSC 1379, Bethesda, MD 20892, USA
[b] Pearson Knowledge Technologies, 4940 Pearl East Circle, Suite 200, Boulder, CO 80301, USA
[c] University of Colorado, Institute for Cognitive Science, Boulder, CO, USA
[d] Harvard Medical School, Boston, MA, USA
[e] The Boston VA Healthcare Services, 940 Belmont Ave, Brockton, MA, USA

ARTICLE INFO

Article history:
Received 13 January 2009
Received in revised form 11 May 2009
Accepted 13 May 2009

Keywords:
Discourse
Semantics
Schizophrenia
Latent semantic analysis

ABSTRACT

Communication disturbances are prevalent in schizophrenia, and since it is a heritable illness these are likely present – albeit in a muted form – in the relatives of patients. Given the time-consuming, and often subjective nature of discourse analysis, these deviances are frequently not assayed in large scale studies. Recent work in computational linguistics and statistical-based semantic analysis has shown the potential and power of automated analysis of communication. We present an automated and objective approach to modeling discourse that detects very subtle deviations between probands, their first-degree relatives and unrelated healthy controls. Although these findings should be regarded as preliminary due to the limitations of the data at our disposal, we present a brief analysis of the models that best differentiate these groups in order to illustrate the utility of the method for future explorations of how language components are differentially affected by familial and illness related issues.

Published by Elsevier Ltd.

1. Introduction

Schizophrenia is widely regarded as a neurodevelopmental disorder in which damage to the brain occurs many years before the illness expresses itself in a florid fashion (Murray & Lewis, 1987;

* Corresponding author.
E-mail address: brita@elvevaag.net (B. Elvevåg).

0911-6044/$ – see front matter Published by Elsevier Ltd.
doi:10.1016/j.jneuroling.2009.05.002

Weinberger, 1987). Therefore it is assumed that even though the actual illness emerges in adulthood, evidence of deficits in brain function is present early in life, albeit in a less dramatic form. Indeed, findings of cognitive weakness being present before illness onset provide strong evidence for abnormal cortical development (David, Malmberg, Brandt, Allebeck, & Lewis, 1997; for a review, see Elvevåg & Weinberger, 2001). Schizophrenia is also considered to be heritable via a polygenic mechanism, such that multiple genes exert relatively small effects that exceed a liability threshold (for a review, see Cannon, 2005). Therefore some similar deficits should be evident in family members, specifically first-degree relatives, although in a muted form.

The bulk of this quest for deficits, in both probands as well as their unaffected relatives, has generally focused on cognitive domains (e.g., working memory, episodic memory, attention) that are considered to be at the very core of the pathology (Bilder et al., 2000; Egan et al., 2001; for reviews, see Elvevåg & Goldberg, 2000; Kuperberg & Heckers, 2000). Since deficits may index genetic liability, they are considered to be candidate intermediate phenotypes for schizophrenia and may be predictive of who develops the actual illness (e.g., see Aukes et al., 2008). Thus, even though schizophrenia is associated with a wide range of symptoms and cognitive deficits (all of which vary in terms of their frequency, predictive validity, specificity, course and amelioration by neuroleptic medication), it is deficits in cognition that have been regarded as the enduring feature of the illness, and have recently become the target for medication and treatment intervention (Kern et al., 2008; Marder & Fenton, 2004; Nuechterlein et al., 2008).

Within this approach, language variables have generally been represented by measures of vocabulary knowledge, reading pronunciation, and counts of the ability to generate as many words beginning with a specific letter or belonging to a specific category in a fixed period of time (e.g., 1 min; for a meta-analysis, see Bokat & Goldberg, 2003). These measures provide very limited windows into language ignoring most aspects of communication, and category fluency for example is more likely tapping into verbal memory than language per se. Despite these rather narrow views of language, there have been some interesting findings. A recent meta-analysis of the cognitive deficits in unaffected first-degree relatives of schizophrenia patients found that of all the cognitive measures examined the largest effect size was with category fluency ($d = .68$; although this effect disappeared with more rigorous inclusion criteria, see Snitz, MacDonald, & Carter, 2006). It is possible that examining the structure within the output of this fluency (i.e., the actual semantic search process itself) may provide useful clues concerning the underlying mechanisms. There is also much literature that adopts a wider approach to examine communicative (rather than linguistics) variables – such as features – in schizophrenia (e.g. see Gernsbacher, Tallent, & Bolliger, 1999 for an overview). Communication analysis is therefore likely to be of enormous value in elucidating the underlying vulnerabilities in this cognitive structure, since communication is a high-level cognitive function that provides a rich and extemporaneous dataset reflecting the state of numerous underlying cognitive processes. The pattern and content of communication provides large amounts of information that can be traced back to individuals' cognitive abilities, knowledge and consequently overall mental state.

An additional advantage of a focus on robust measurement tools of cognition is that they can be used to more specifically define and explore the underlying psychopathology of the disorder and also focus specifically on aspects peculiar to schizophrenia, such as disorganized thinking, as evidenced by disorganized speech. Although it may be argued that 'unconventional' use of language is simply a characteristic of the acute psychotic state and subsides when the psychosis does, studies show that even in the stable state, several characteristics of language processing are not 'conventional' in people with schizophrenia (e.g., Li, Branch, Ardekani, et al., 2007; Li, Branch, Bertisch, et al., 2007; Sommer, Ramsey, & Kahn, 2001; Sommer, Ramsey, Mandl, & Kahn, 2003). However, if 'unconventional' use of language are trait abnormalities the assumption is that communication, as a complex combination of cognitive processes, may account for some of the genetic burden if it can be usefully assayed. Indeed, there is a strong theoretical rationale for analyzing language samples from individuals at high genetic risk for schizophrenia, as the neural pathways for language processing are likely related to the underlying pathophysiology of the disorder (DeLisi, 2001; Li, Branch, Ardekani, et al., 2007; Li, Branch, Bertisch, et al., 2007).

Since brain pathology is already detectable by the time of first episode (and probably progressing in the prodromal phase before symptoms appear), one goal would be to detect a range of subtle discourse