Reflections on measuring disordered thoughts as expressed via language

Brita Elvevåg

Department of Clinical Medicine, University of Tromsø - the Arctic University of Norway, Tromsø, Norway

ARTICLE INFO

ABSTRACT

Thought disorder, as inferred from disorganized and incoherent speech, is an important part of the clinical presentation in schizophrenia. Traditional measurement approaches essentially count occurrences of certain speech events which may have restricted their usefulness. Applying speech technologies in assessment can help *automate* traditional clinical rating tasks and thereby complement the process. Adopting these computational approaches affords clinical translational opportunities to *enhance* the traditional assessment by applying such methods remotely and scoring various parts of the assessment automatically. Further, digital measures of language may help detect subtle clinically significant signs and thus potentially *disrupt* the usual manner by which things are conducted. If proven beneficial to patient care, methods where patients' voice are the primary data source could become core components of future clinical decision support systems that improve risk assessment. However, even if it is possible to measure thought disorder in a sensitive, reliable and efficient manner, there remain many challenges to then translate into a clinically implementable tool that can contribute towards providing better care. Indeed, embracing technology - notably artificial intelligence - requires vigorous standards for reporting underlying assumptions so as to ensure a trustworthy and ethical clinical science.

1. Ordering thoughts about disordered thinking

Disturbances in the organization and coherence of thought that characterize the formal thought disorder present in many patients with schizophrenia are fascinating scientifically and clinically. Regardless of whether thought disorder is pathognomic of schizophrenia, as a symptom it comprises a major part of the phenomenology and is important in the diagnosis and treatment process. Given the numerous cognitive deficits that present in schizophrenia and in those with thought disorder, it seems logical to assume that thought disorder could be a consequence of some of these processes such as for example due to a problem in working memory, attention, language and / or semantic memory. However, right at the beginning of my postdoctoral fellowship it was clear that: 'Although each psychological construct can account for some of the data, establishing a primary cognitive impairment responsible for thought disorder is not forthcoming, perhaps because the underlying pathology is multidimensional' (Elvevåg and Goldberg, 1997). A problem in semantic organization seemed potentially promising, yet despite elaborate cognitive scientific attempts to examine at a detailed level, the findings in this cognitive process were probably most surprising. For example, to evaluate knowledge about natural categories and its organization, we examined the relationships between different taxonomic levels. Specifically, participants were asked to generate exemplars such as 'dog' to the superordinate 'animal' and subsequently rated the typicality to the superordinate 'animal', and we then modeled the relationship to the category level 'mammal'. Importantly, the mathematical instantiation model predicted the statistical taxonomic relationships in a comparable manner in patients with schizophrenia and healthy participants, thus refuting the simple notion that the unusual content and structure in patients' language reflects a disturbance in their actual knowledge contained in semantic concepts (and in the words used to refer to these concepts) (Elvevåg et al., 2005). Clearly the complex and dynamical processes underlying free speech are considerably more difficult than required in this constrained language task. However, what was especially unsettling in all of this was how disordered thought was actually measured. The measurement process necessitates a fairly substantial (e.g., 45 min) interview with a patient which is obviously demanding in many ways, notably the time and effort requested from a patient - who by definition is ill - as well as the time required by an interviewer. Additionally, there is a lot of scope for subjectivity to creep in (e.g., because of distraction) and it is in essence a matter of counting occurrences of various aspects of speech. Put simply, it is an interview that obviously cannot be conducted very frequently (e.g., daily is impossible) despite thought disorder being something that can change very rapidly (e.g., on the order of a few days). Yet it is frequent measurement that is necessary because if measured appropriately could be useful in terms of understanding and monitoring the treatment process.

Therefore, it seemed necessary to think differently, namely look for a technological alternative that could both do it differently and hopefully require a shorter sampling of speech. My reflections below are structured as follows: Technology can help (i) *automate* what we do, (ii) *enhance* what we do, and (iii) completely *disrupt* the usual manner by which we do things. In the case of psychiatry leveraging speech technologies, there is the potential to check off all three of these points, both in a manner that is expected but hopefully also in ways we have yet to discover.

2. Listening to stories: symptoms or signs?

Still in the 21st century, the patient's vocal self-presentation as elicited in a clinical interview remains central in the psychiatric diagnosis process and in monitoring treatment effectiveness. Indeed, getting patients to tell their personal story by them recalling information that clinicians consider useful is central in any patient-clinician interaction. Although such deeply personal story telling generates phenomenological information about symptoms, a quite different and formalized reductionist approach to story-telling might reveal critical clues about mental health that provide information regarding medical signs.

Rating symptoms from the speech of patients in a reliable fashion can be really difficult. At the beginning of my postdoctoral fellowship (1996), when being trained on using a symptom rating scale to characterize patients' speech in a clinical interview, my poor skills were contributing to some embarrassing inter-rater reliability scores in the trainee group. Therefore, I started to day-dream about the day when a machine would be able to conduct these clinical ratings better than I could (i.e., (i) technology *automating* what experts do), and maybe do so in a manner that was more objective (i.e., (ii) technology *enhancing* what experts do). Also, acutely aware of the inequity in medical care and the disproportionate number of patients who simply cannot meet with clinicians as often as necessary, I longed for the day when my future virtual helper that provided an automated second opinion could stand in for me and do so via the phone so as to reach more patients more frequently (i.e., (iii) technology *disrupting* the traditional status quo). These three points in my day-dream summarize much of my psychiatric research and scientific goals.

3. New stories to tell: listening with computational ears

As a student I became intrigued by the computational models emerging in behavioral science that were leveraging parallel distributed processing models, and the emerging field of computational psychiatry (notably the pioneering work of Eric Chen (Chen, 1994; Chen 1995), Jonathan Cohen and David Servan-Schreiber (Cohen and Servan-Schreiber, 1992), and Ralph Hoffman (Hoffman, 1987)). The experimental frameworks linking symptoms to underlying neurocognition that this genre of computational research generated were novel as were the metaphors, but importantly the models were testable and thus refutable. The very idea of using computer simulations necessitates computationally explicit assumptions, as models expressed via computer programs require implicit assumptions to be detailed enough to implement.

This attention to underlying detail was useful in my experimental psychological approach to deconstruct the cognitive substrates of schizophrenia and in refining the cognitive phenotypes by developing neurocognitive assays that better characterized the fundamental operations closest to the neurobiology (see e.g., Elvevåg and Goldberg, 2000; Elvevåg and Weinberger, 2001). My interest was in potential cognitive mechanisms associated with patients' thought disorder as expressed via speech. There was a lot of skepticism at the time that language would be a useful metric to focus on in terms of being a useful phenotype. However, I was deeply impressed by the boldness of the ideas relating to language that were championed by Tim Crow and by Lynn DeLisi regarding differences (e.g., reductions or reversals) in structural cerebral asymmetries being related to the pathogenesis of schizophrenia (DeLisi

et al., 1997; DeLisi 2001). Tim Crow had also boldly proposed that a gene associated with the evolution of human language and cerebral specialization might actually cause schizophrenia (Crow et al., 1989). Of course 'asymmetric pathology is not necessarily pathology of asymmetry' and 'even if the asymmetry data turn out not to be epiphenomena, there is no obvious explanation for why an abnormality in language cortex asymmetry would translate into psychosis in later life, rather than simply a developmental language disorder' (quotes from Elvevåg and Weinberger, 1997). Nonetheless at the core of these bold theories regarding anomalous lateralization were reports of more non-right-handedness in schizophrenia. However, our examination of handedness, neuropsychological performance and brain asymmetries with structural magnetic resonance imaging data to compare patients with schizophrenia, their unaffected siblings and unrelated healthy controls did not find evidence for non-right-handedness being a schizophrenia risk-associated heritable phenotype (Deep-Soboslay et al., 2010).

Late in 1996, I met Peter Foltz who was working on rating student essays using computational semantics (Foltz, 1996). Given my poor rating skills in clinical interviews, I was very eager to adopt a similar approach with the speech of patients. We started immediately the first of our many studies using computational language processing methods to analyze patients' speech (for a reflection of our 25 years of hindsight, see Foltz et al., 2022). I was very confident that these methods would at the very least outperform me! We applied natural language processing methods to the language generated from clinical interviews and also from more formal tasks such as word association, verbal fluency and notably story recollection. The purpose was to use these language samples to provide a second opinion about illness severity and to further our understanding of the underlying neurocognitive mechanisms of disordered thinking. As expected, we could derive a range of useful indices such as speech coherence, tangentiality and measures of the amount of relevant content in responses and importantly these correlated with expert clinician assessments. This research - like the work on rating student essays - was met with considerable skepticism and significant hostility and it was extremely hard to publish. Nonetheless this new approach to complement and improve psychiatric assessment successfully differentiated the language in patients with schizophrenia from healthy controls (Elvevåg et al., 2007), first-degree relatives (Elvevåg et al., 2010), and those at high risk of psychosis (Rosenstein et al., 2015).

The bigger picture was that this approach of using natural language technologies in psychiatric assessment was *technologically disruptive* in that it enabled scaling up and thus laid the foundation for remote monitoring longitudinally using language where complete automation would enable this on a large scale (Chandler et al., 2019, 2020a; Holmlund et al., 2019). It also provided a different way of thinking about operationalizing, measuring and analyzing discourse in order to understand disordered language production. However, to the extent that tests constrain a construct, it was very clear that the field needed a broader way of defining language rather than relying on existing (old and constrained) tests defining language that has resulted in a grossly inadequate view of language abilities or how we might measure them (Elvevåg et al., 2016; Elvevåg et al., 2017).

4. The future of telling stories: new ways of doing things

At the time of writing, an industry of staggering growth is the healthcare chat-bots that are encouraging patients to share their personal stories with the inanimate chat-bot, and these conversations are analyzed with state of the art natural language processing tools. Indeed, stories are fundamental to our human experience and provide an effective way for us to organize information, and thus evaluating how stories are recalled provides critical information about mental health and memory function. Obviously there are very many different cognitive steps and processes that occur between the questions being asked by a clinician and a patient responding. Traditional neuropsychology has

adopted a reductionist approach to measure these separate behavioral constructs such as verbal memory or attention. However, a future alternative might be that similar (but not identical) constructs are derived from the storytelling process by aligning these constructs to the story recall derived computational features (Foltz et al., 2022). In principle, this would make it possible to derive some useful medical signs from this 'mental blood test' comprised of simply a few minutes of speech and to do so in real time (for an animation of a possible near future method of measuring coherence in an interview, see https://ars. els-cdn.com/content/image/1-s2.0-S0920996422003620-mmc1.gif - from Holmlund et al., 2022). However, to transform this science fiction notion to reality requires that story-telling as a speech elicitation task be designed for this purpose and certainly the components measured would need to be theoretically and clinically motivated. A variety of comparable stories would enable frequent administration, such that this could be evaluated longitudinally and remotely. The speech would be scored automatically and deep semantic themes evaluated (Chandler, 2019, 2021; Diaz-Asper et al., 2022; Holmlund et al., 2020). Such frequent and varied testing would enable earlier and less granular insight into patients' evolving cognitive and mental health. However, to really reap the full benefits of this a new psychometrics is needed, where dynamics is central to understanding the individual. This framework will be necessary when using longitudinal data to understand how temporal dynamical changes relate to cognitive and mental states, but also be critical for real-time modeling to prospectively predict future mental and cognitive states.

5. Conclusion

The hope or assumption is that if one has a reliable way to assay signs that it might be possible to use this to capture the warning signs *before* something bad happens. There are many parallels in nature. Consider the west African jungle, where over time one can become fairly skilled at learning to listen to the 'roar' of the jungle, the change in types of sounds, intensity and overall 'melody' in the air, as seemingly the animals 'sense' a pending storm, giving those who have managed to listen properly, ample time to seek suitable shelter. The warning signals are there, but we just need to listen to the correct ones. Could this be done in psychiatry? It probably can be and currently there is no shortage of attempts to leverage smart devices and remotely monitor outpatients. However, like weather forecasting (e.g., Bauer et al., 2015), knowing *what* combination of signals to focus on for the specific event (e.g., clinical state change versus memory decline) and *when* it will translate into something (e.g., in a few hours, tomorrow, next week, next month, next year) remains a remarkably complex problem (Holmlund et al., 2021). Further, modeling time at the various relevant time scales will be essential both to understand the micro details of the dynamic nature of behavior including language, as well as accurately leveraging this temporal information to detect pending clinical events (and avoiding false positives). In doing this it will be critical that the entire process nurtures trustworthiness, namely from design through to methods employed, data details and analysis techniques underlying the modeling (Chandler et al., 2020b). However, the complex ethics and legality of actually alerting (or not) entities (e.g., individuals, employers, clinicians, families) that there *might* be some clinical event about to happen *sometime* in the future may well turn out to be considerably more challenging than imagined.

Declaration of Competing Interest

The author does not have any conflicts of interests to disclose.

References

Bauer, P., Thorpe, A., Brunet, G., 2015. The quiet revolution of numerical weather prediction. Nature 525 (7567), 47–55. https://doi.org/10.1038/nature14956.

Chandler, C., Foltz, P.W., Cheng, J., Bernstein, J.C., Rosenfeld, E.P., Cohen, A.S., Holmlund, T.B., Elvevåg, B., 2019. Overcoming the bottleneck in traditional assessments of verbal memory: modeling human ratings and classifying clinical group membership. In: Niederhoffer, K., Hollingshead, K., Resnik, P., Resnik, R., Loveys, K. (Eds.), Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology. Minneapolis, Minnesota, USA, pp. 137–147. https://doi.org/10.18653/v1/W19-3016.

Chandler, C., Foltz, P.W., Cohen, A.S., Holmlund, T.B., Cheng, J., Bernstein, J.C., Rosenfeld, E.P., Elvevåg, B., 2020a. Machine learning for longitudinal applications of neuropsychological testing. Intelligence-Based Med. 1-2, 100006 https://doi.org/10.1016/j.ibmed.2020.100006.

Chandler, C., Foltz, P.W., Elvevåg, B., 2020b. Using machine learning in psychiatry: the need to establish a framework that nurtures trustworthiness. Schizophr. Bull. 46, 11–14. https://doi.org/10.1093/schbul/sbz105.

Chandler, C., Holmlund, T.B., Foltz, P.W., Cohen, A.S., Elvevåg, B., 2021. Extending the usefulness of the verbal memory test: the promise of machine learning. Psychiatry Res. 297, 113743 https://doi.org/10.1016/j.psychres.2021.113743.

Chen, E.Y.H., 1994. A neural network model of cortical information processing in schizophrenia. I: interaction between biological and social factors in symptom formation. Can. J. Psychiatry 39 (8), 362–367. https://doi.org/10.1177/070674379403900803.

Chen, E.Y.H., 1995. A neural network model of cortical information processing in schizophrenia. II: role of hippocampal-cortical interaction: a review and a model. Can. J. Psychiatry 40 (1), 21–26. https://doi.org/10.1177/070674379504000107.

Cohen, J.D., Servan-Schreiber, D., 1992. Context, cortex and dopamine: a connectionist approach to behaviour and biology in schizophrenia. Psychol. Rev. 99, 45–77. https://doi.org/10.1037/0033-295x.99.1.45.

Crow, T.J., Ball, J., Bloom, S.R., Brown, R., Bruton, C.J., Colter, N., Frith, C.D.-., Johnstone, E.C.-., Owens, D.G., Roberts, G.W., 1989. Schizophrenia as an anomaly of development of cerebral asymmetry. A postmortem study and a proposal concerning the genetic basis of the disease. Arch. Gen. Psychiatry 46 (12), 1145–1150. https://doi.org/10.1001/archpsyc.1989.01810120087013.

Deep-Soboslay, A., Hyde, T.M., Callicott, J.P., Lener, M.S., Verchinski, B.A., Apud, J.A., Weinberger, D.R., Elvevåg, B., 2010. Handedness, heritability, neurocognition and brain asymmetry in schizophrenia. Brain 133 (10), 3113–3122. https://doi.org/10.1093/brain/awq160.

DeLisi, L.E., Sakuma, M., Kushner, M., Finer, D.L., Hoff, A.L., Crow, T.J., 1997. Anomalous cerebral asymmetry and language processing in schizophrenia. Schizophr. Bull. 23 (2), 255–271. https://doi.org/10.1093/schbul/23.2.255. Erratum in: *Schizophr Bull* 1997;23(3):536.

DeLisi, L.,.E., 2001. Speech disorder in schizophrenia: review of the literature and exploration of its relation to the uniquely human capacity for language. Schizophr. Bull. 27 (3), 481–496. https://doi.org/10.1093/oxfordjournals.schbul.a006889.

Diaz-Asper, C., Chandler, C., Turner, R.S., Reynolds, B., Elvevåg, B., 2022. Increasing access to cognitive screening in the elderly: applying natural language processing methods to speech collected over the telephone. Cortex 156, 26–38. https://doi.org/10.1016/j.cortex.2022.08.005.

Elvevåg, B., Goldberg, T.E., 1997. Formal thought disorder and semantic memory in schizophrenia. CNS Spectr 2 (8), 15–25. https://doi.org/10.1017/S1092852900005009.

Elvevåg, B., Weinberger, D.R., 1997. Schizophrenia and the right shift gene hypothesis: some comments on the devil in the details. Cogn Neuropsychiatry 2 (3), 221–225.

Elvevåg, B., Goldberg, T.E., 2000. Cognitive impairment in schizophrenia is the core of the disorder. Crit Rev Neurobiol 14 (1), 1–21. PMID: 11253953.

Elvevåg, B., Weinberger, D.R., 2001. Neuropsychology in context of the neurodevelopmental model of schizophrenia. In: Nelson, C.A., Luciana, M. (Eds.), Handbook of Developmental Cognitive Neuroscience. MIT Press, Cambridge, Mass, pp. 577–595.

Elvevåg, B., Heit, E., Storms, G., Goldberg, T., 2005. Category content and structure in schizophrenia: an evaluation using the instantiation principle. Neuropsychology 19 (3), 371–380. https://doi.org/10.1037/0894-4105.19.3.371.

Elvevåg, B., Foltz, P.W., Weinberger, D.R., Goldberg, T.E., 2007. Quantifying incoherence in speech: an automated methodology and novel application to schizophrenia. Schizophr. Res. 93 (1–3), 304–316. https://doi.org/10.1016/j.schres.2007.03.001.

Elvevåg, B., Foltz, P.F., Rosenstein, M., DeLisi, L., 2010. An automated method to analyze language use in patients with schizophrenia and their first-degree relatives. J Neurolinguistics 23, 270–284. https://doi.org/10.1016/j.jneuroling.2009.05.002.

Elvevåg, B., Cohen, A.S., Wolters, M.K., Whalley, H.C., Gountouna, V.E., Kuznetsova, K. A., Watson, A.R., Nicodemus, K.K., 2016. An examination of the language construct in NIMH's Research Domain Criteria: time for reconceptualisation! Am. J. Med. Genetics Part B: Neuropsychiatric Genetics 6 (171), 909–919. https://doi.org/10.1002/ajmg.b.32438.

Elvevåg, B., Foltz, P.W., Rosenstein, M., Ferrer-I-Cancho, R., De Deyne, S., Mizraji, E., Cohen, A., 2017. Thoughts about disordered thinking: measuring and quantifying the laws of order and disorder. Schizophr. Bull. 43 (3), 509–513. https://doi.org/10.1093/schbul/sbx040.

Foltz, P.W., 1996. Latent Semantic Analysis for text-based research. Behav. Res. Methods, Instrum. Comput. 28 (2), 197–202. https://doi.org/10.3758/BF03204765.

Foltz, P.W., Chandler, C., Diaz-Asper, C., Cohen, A.S., Rodriguez, Z., Holmlund, T.B., Elvevåg, B., 2022. Reflections on the nature of measurement in language-based automated assessments of patients' mental state and cognitive function. Schizophr Res. https://doi.org/10.1016/j.schres.2022.07.011. S0920-9964(22)00283-3Epub ahead of print. PMID: 36153250.

Hoffman, R.E., 1987. Computer simulations of neural information processing and the schizophrenia-mania dichotomy. Arch. Gen. Psychiatry 44 (2), 178–188. https://doi.org/10.1001/archpsyc.1987.01800140090014.

Holmlund, T.B., Foltz, P.W., Cohen, A.S., Johansen, H.D., Sigurdsen, R., Fugelli, P., Bergsager, D., Cheng, J., Bernstein, J., Rosenfeld, E., Elvevåg, B., 2019. Moving psychological assessment out of the controlled laboratory setting and into the hands of the individual: practical challenges. Psychol. Assess. 31 (3), 292–303. https://doi.org/10.1037/pas0000647.

Holmlund, T.B., Chandler, C., Foltz, P.W., Cohen, A.S., Cheng, J., Bernstein, J.C., Rosenfeld, E.P., Elvevåg, B., 2020. Applying speech technologies to assess verbal memory. npj Digital Med. 3, 33. https://doi.org/10.1038/s41746-020-0241-7.

Holmlund, T.,.B., Diaz-Asper, C., Elvevåg, B., 2021. The reality of doing things with (thousands of) words in applied research and clinical settings: a commentary on Clarke et al. (2020). Cortex 136, 150–156. https://doi.org/10.1016/j.cortex.2020.08.024.

Holmlund, T.B., Chandler, C., Foltz, P.W., Diaz-Asper, C., Cohen, A.S., Rodriguez, Z., Elvevåg, B., 2022 Nov 10. Towards a temporospatial framework for measurements of disorganization in speech using semantic vectors. Schizophr Res. https://doi.org/10.1016/j.schres.2022.09.020. S0920-9964(22)00362-0Epub ahead of print. PMID: 36372683.

Rosenstein, M., Foltz, P.W., DeLisi, L.E., Elvevåg, B., 2015. Language as a biomarker in those at high-risk for psychosis. Schizophr. Res. 165, 249–250. https://doi.org/10.1016/j.schres.2015.04.023.