

nied by other initiatives such as public education and legislation. As with most complex problems, the solution needs to be strategic, coordinated and multi-pronged. More research is also vital, as consensus on any new name should ideally be derived from a large, diverse sample of all relevant stakeholders and a rigorous scientific consensus. It is particularly critical to continue to include the voices of people who live with the condition, who are often marginalized and suffer inequities, a point cogently and eloquently illustrated by a recent paper in this journal<sup>9</sup> describing the lived experience of psychosis.

Words matter. If a name change can even be part of what leads to improved lives for people with the condition, then isn't it worth it? Why keep a name that the majority of people with the condition are not comfortable with, that they feel is stigmatizing and discriminates against them, and that dissuades them from seeking out care? Isn't that reason enough?

What's in a name? Names shift to reflect transformation, and new names catalyze change. As E. Dickinson wrote, "I know nothing in the world that has as much power as a word".

Raquelle I. Mesholam-Gately<sup>1,2</sup>, Dan Johnston<sup>3</sup>, Matcheri S. Keshavan<sup>1,2</sup>  
<sup>1</sup>Department of Psychiatry, Harvard Medical School, Boston, MA, USA; <sup>2</sup>Massachusetts Mental Health Center; Public Psychiatry Division of the Beth Israel Deaconess Medical Cen-

ter, Boston, MA, USA; <sup>3</sup>Consumer Advisory Board, Massachusetts Mental Health Center, Public Psychiatry Division, Beth Israel Deaconess Medical Center, Boston, MA, USA

The authors thank all members of the Consumer Advisory Board of Massachusetts Mental Health Center and other contributors to the Renaming Schizophrenia Survey Project for their ideas, inspiration, and review of this letter.

1. Keshavan MS, Nasrallah HA, Tandon R. *Schizophr Res* 2011;127:3-13.
2. Brown K. "Schizophrenia" still carries a stigma. Will changing the name help? *New York Times*, December 20, 2021.
3. Metz J. The protest psychosis: how schizophrenia became a black disease. Boston: Beacon Press, 2009.
4. Yamaguchi S, Mizuno M, Ojio Y et al. *Psychiatry Clin Neurosci* 2017;71:347-62.
5. Lasalvia A, Vita A, Bellomo A et al. *Schizophr Res* 2021;228:502-9.
6. Mesholam-Gately RI, Varca N, Spitzer C et al. *Schizophr Res* 2021;238:152-60.
7. Saks ER. *Schizophr Res* 2022;242:121-2.
8. Keshavan MS, Davis B, Friedman-Yakoobian M et al. *Schizophr Res* 2022;239:92-4.
9. Fusar-Poli P, Estradé A., Stanghellini G et al. *World Psychiatry* 2022;21:168-88.

DOI:10.1002/wps.21033

## Are language features associated with psychosis risk universal? A study in Mandarin-speaking youths at clinical high risk for psychosis

Natural language processing (NLP) analyses have shown decreased coherence (tangentiality, derailment) and complexity (poverty of content) in schizophrenia and in clinical high risk (CHR) states for psychosis. We reported previously in this journal<sup>1</sup> that an NLP machine learning classifier, which included measures of coherence and complexity, predicted psychosis onset in two independent English-speaking CHR samples. Moreover, reduced complexity has been associated with increased pauses and negative symptoms in at-risk youths<sup>2</sup>.

Multiple recent NLP studies in schizophrenia and CHR cohorts, using different methods, have largely found this same pattern of disturbance in the structure of language and speech<sup>3</sup>. Most of these studies have been conducted in English, with notable exceptions including Dutch, Portuguese and Spanish<sup>4</sup>. It remains unknown, however, whether NLP findings obtained in English or other Indo-European languages would generalize to less similar languages, such as Mandarin, which has very different grammatical and prosodic conventions.

This study included 20 help-seeking CHR youth and 25 healthy controls who were recruited as part of the US National Institute of Mental Health (NIMH)-funded Shanghai-At-Risk for Psychosis (SHARP) study at the Shanghai Mental Health Center, where institutional review board approval was obtained. Caseness and symptoms were determined using the Structured Interview for

Psychosis-Risk Syndromes (SIPS)<sup>5</sup>. Subjects were Han Chinese and spoke Mandarin fluently, and they provided informed consent. Sex distribution was similar between CHR subjects and controls (55% vs. 48% female), but CHR subjects were younger (19.6±6.4 vs. 24.9±1.9 years) and had less education (11.4±4.0 vs. 16.7±1.4 years).

Interviews were approximately 30 min in length, and were based on qualitative methods previously described<sup>6</sup>. They were transcribed verbatim in Mandarin and translated into English using Google Translate, with verification by bilingual researchers. Audio recordings were diarized (segmented by speaker using time stamps from transcription) so that acoustic analyses could be done of subjects' speech.

NLP features analyzed for both English and Mandarin included coherence, complexity, and sentiment (i.e., emotional valence – positive, negative, neutral), as reported previously<sup>1,7</sup>. For English NLP only, sentiment also included anger, fear, sadness, joy and disgust; frequency of wh-words (e.g., "which") was also assessed. For Mandarin NLP only, frequency of measure words, possessives, and localizers (e.g., *gongzuo-shang*, "during work"; or *liangge-ren-zhijian*, "between two people") was also calculated<sup>8</sup>. Acoustic features analyzed in Mandarin included those characteristic of schizophrenia or CHR states among English-speaking subjects, including abnormal pauses, flat intonation, voice breaks, and

pitch variation<sup>7</sup>.

All features were corrected for age and education by applying regression coefficients from healthy controls, and highly correlated features were removed from analysis. Machine learning classification was done using random forest and support vector machines (SVM) for Mandarin NLP, English NLP, and acoustics, with each experiment repeated 20 times, identifying the top five features of each model. Associations between linguistic features (cross-language analysis) and with symptoms (symptom inference) were also tested (see also supplementary information).

Each of the three SVM machine learning classifiers showed high accuracy in discriminating spoken language in CHR subjects from that of healthy controls: English-specific NLP (95%), Mandarin-specific NLP (94%), and acoustic analysis (88%), with similar results for random forest. Top features for the English-specific NLP machine learning were wh-word and noun use (greater in CHR), and coherence, adjective use and adverb use (all less in CHR). Top features for the Mandarin-specific NLP machine learning were localizer use (greater in CHR), and positive sentiment, two metrics of coherence, and adjective use (all decreased in CHR). Of note, features common to the NLP machine learning for both languages were highly correlated, specifically coherence ( $r=.70$ ) and adjective use ( $r=.60$ ).

For acoustics, the top features in the machine learning classifier were two pause metrics, and three indices of acoustic quality: chroma #11 (timbre/quality), bandwidth formant #1 (dysphonia/hoarseness), and spectral spread (energy – decibels/frequency). Of note, only acoustic features were significantly associated with symptoms (negative:  $r=0.69$ ,  $p=8E-4$ , positive:  $r=0.49$ ,  $p=3E-2$ ) (see also supplementary information).

Several important findings emerge from this proof-of-principle study. First, in Mandarin, spoken language can differentiate CHR subjects from healthy controls with high accuracy, using either linguistic or acoustic features. Second, the application of English-specific NLP to transcripts translated from Mandarin has utility, as there was comparable accuracy for both the English-specific and Mandarin-specific NLP. Further, there was overlap in top features in the two NLP classifiers, specifically decreases in adjective use and coherence, with both of these features highly correlated across the two languages, suggesting that these key metrics survive translation. Nonetheless, the application of Mandarin-specific NLP allowed the identification of a key linguistic feature that would not be captured otherwise – the increased use of localizers among CHR subjects – which may reflect concreteness or increased use of idioms; this is a new finding that merits replication and further investigation. Finally, the acoustic classifier, in addition to having high accuracy, identified features similar to those found in English-speaking CHR and schizophrenia cohorts, including abnormal pause behavior, and indices of voice quality and energy. As in prior studies, acoustic features were associated with symptoms, in particular negative symptoms.

This study is the first to use natural language processing and acoustic analyses to characterize spoken language among native

Mandarin speakers in China identified as at clinical high risk for psychosis. Our findings support the idea that there may be universal features of spoken language disturbance across psychosis and its risk states, particularly in respect to reduced coherence, but also word usage and pause behavior that may index reduced complexity. Yet, our study also shows that there are language-specific features characteristic of psychosis risk, suggesting that it is essential to also analyze spoken language using language-specific NLP methods.

This is a small proof-of-principle study with the potential confounds of age and education, and none of the classifiers generated were cross-validated in a second cohort. Therefore, these findings should be investigated and replicated in a larger cohort of Mandarin-speaking CHR subjects and healthy controls who are more similar in demographics.

More broadly, future studies should include a similar heuristic of using both English-based and language-specific NLP approaches, as well as acoustic analyses, to assess spoken language in CHR cohorts (e.g., English, Mandarin, Cantonese, Korean, Spanish, German, Portuguese, Danish, French, Italian) from around the world, as is planned for the Accelerating Medicines Partnership<sup>®</sup> Program – Schizophrenia, to determine both universal and language-specific features of language disturbance characteristic of clinical risk for psychosis.

Carla Agurto<sup>1</sup>, Raquel Norel<sup>1</sup>, Bo Wen<sup>1</sup>, Yanyan Wei<sup>2</sup>, Dan Zhang<sup>2</sup>, Zarina Bilgrami<sup>3</sup>, Xiaolu Hsi<sup>4</sup>, Tianhong Zhang<sup>2</sup>, Ofer Pasternak<sup>5</sup>, Huijun Li<sup>6</sup>, Matcheri Keshavan<sup>4</sup>, Larry J. Seidman<sup>4</sup>, Susan Whitfield-Gabrieli<sup>7</sup>, Martha E. Shenton<sup>5,8</sup>, Margaret A. Niznikiewicz<sup>9</sup>, Jijun Wang<sup>2</sup>, Guillermo Cecchi<sup>1</sup>, Cheryl Corcoran<sup>10</sup>, William S. Stone<sup>4</sup>

<sup>1</sup>IBM TJ. Watson Research Center, Yorktown Heights, NY, USA; <sup>2</sup>Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China; <sup>3</sup>Department of Psychology, Emory University, Atlanta, GA, USA; <sup>4</sup>Department of Psychiatry, Beth Israel Deaconess Medical Center; Harvard Medical School, Boston, MA, USA; <sup>5</sup>Department of Psychiatry, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; <sup>6</sup>Department of Psychology, Florida A&M University, Tallahassee, FL, USA; <sup>7</sup>Department of Psychology, Northeastern University, Boston, MA, USA; <sup>8</sup>Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; <sup>9</sup>Department of Psychiatry, VA Boston Healthcare System, Brockton, MA, USA; <sup>10</sup>Icahn School of Medicine at Mount Sinai, New York, NY, USA

The authors would like to thank the study participants at the Shanghai Mental Health Center. C. Agurto and R. Norel contributed equally to the work. Supplementary information on this study can be found at <https://wenboown.github.io/SHARP-NLP-Letter-to-WP/>.

1. Corcoran CM, Carrillo F, Fernandez-Slezak D et al. *World Psychiatry* 2018; 17:67-75.
2. Stanislawski ER, Bilgrami ZR, Sarac C et al. *NPJ Schizophr* 2021;7:3.
3. Corcoran CM, Cecchi G. *Biol Psychiatry Cogn Neurosci Neuroimaging* 2020; 5:770-9.
4. Corcoran CM, Mittal VA, Bearden CE et al. *Schizophr Res* 2020;226:158-66.
5. McGlashan TH, Walsh BC, Woods SW. *Structured Interview for Psychosis-Risk Syndromes*. New Haven: PRIME Research Clinic, Yale School of Medicine, 2017.
6. Ben-David S, Birnbaum ML, Eilenberg ME et al. *Psychiatr Serv* 2014;65:1499-501.
7. Agurto C, Cecchi GA, Norel R et al. *Neuropsychopharmacology* 2020;45:823-32.
8. Chappell H, Peyraube A. In: Xu D (ed). *Space in languages of China: cross-linguistic, synchronic and diachronic perspectives*. Dordrecht: Springer, 2008: 15-37.

DOI:10.1002/wps.21045