

Quantization of speech disorganization for PTSD and speech disorders detection

Samuel Boccara¹, Salma Mesmoudi^{2,3}, Jacques Dayan⁴, Robin Quillivic^{2,5}

¹Ecole Nationale Supérieure de l'Electronique et de ses Applications (ENSEA); Ministère de l'Enseignement Supérieur et de la Recherche Scientifique – samuel.boccara@ensea.fr

²Institut des Systèmes Complexes - Paris Ile-de-France (ISC-PIF); École normale supérieure - Cachan, Université Paris 1 Panthéon-Sorbonne, Sorbonne Université, Centre National de la Recherche Scientifique – salma.mesmoudi@iscpif.fr

³Centre européen de sociologie et de science politique (CESSP) ; Université Paris 1 Panthéon-Sorbonne, École des Hautes Études en Sciences Sociales, Centre National de la Recherche Scientifique, UMR8209

⁴Neuropsychologie et imagerie de la mémoire humaine (NIMH), Université de Caen Normandie, GIP Cyceron, Institut National de la Santé et de la Recherche Médicale, Ecole Pratique des Hautes Etudes – Jacques.dayan@chu-rennes.fr

⁵École pratique des hautes études (EPHE), Université Paris sciences et lettres – robin.quillivic@ephe.sorbonne.fr

Abstract

Psycholinguistic literature shows that Post-Traumatic Stress Disorder (PTSD) and its specific criteria as stated in the DSM-5 affect language. We aim to analyze these effects using automated methods while keeping interpretability, to quantify them and their importance. However, automatic PTSD detection is an open problem, with NLP-based approaches being even rarer. With psychiatrists often having interviews with the patient, the problem itself of detecting PTSD from a single discourse without interaction is a complex one and gold standard tests have only 80% accuracy in doing so. In this study, we propose an approach, focusing on speech disorganization effects only and their association with PTSD. Our method, based on sentence similarity, used state-of-the-art models to underline speech organization in transcribed discourse. Using such representation allowed us to create features inferred from the way psychiatrists read texts. On the 13-November Cohorts, our method reached accuracy comparable to human standards for our dataset, both in PTSD prediction and Criterion D (negative alterations in cognitions and mood) detection. According to the literature, criterion D is the most related to speech organization. We also added a second contribution to our work: a method extracting parts of texts that contain the speech disorder phenomenon.

Keywords: NLP; Psycholinguistics, PTSD; disorganization of speech.