



**HAL**  
open science

# Quantization of speech disorganization for PTSD and speech disorders detection

Samuel Boccara, Salma Mesmoudi, Jacques Dayan, Robin Quillivic

► **To cite this version:**

Samuel Boccara, Salma Mesmoudi, Jacques Dayan, Robin Quillivic. Quantization of speech disorganization for PTSD and speech disorders detection. JADT 2024, LesLa, Jun 2024, Bruxelles, Belgium. pp.ISBN : 978-2-39061-471-5. hal-04627800v2

**HAL Id: hal-04627800**

**<https://hal.science/hal-04627800v2>**

Submitted on 3 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Quantization of speech disorganization for PTSD and speech disorders detection

Samuel Boccara<sup>1</sup>, Salma Mesmoudi<sup>2,3</sup>, Jacques Dayan<sup>4</sup>, Robin Quillivic<sup>2,5</sup>

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

<sup>2</sup>Institut 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

<sup>3</sup>Centre 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

<sup>4</sup>Neuropsychologie 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

<sup>5</sup>École 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.

## 1. Introduction

PTSD is a frequent endemic mental health condition that might develop after a person experiences or witnesses a traumatic event (Criterion A). Symptoms of PTSD can be grouped into 5 main categories according to DSM-5 (American Psychiatric Association, 2013.): re-experiencing the traumatic event (criterion B), avoidance of reminders of the event (criterion C), negative changes in cognition and mood (criterion D), changes in physical and emotional reactions (criterion E), these symptoms must create distress or functional impairment in a person's life (criterion G) and last for more than a month (Criterion F). Psycholinguistic literature and namely the SPLIT study (Auxéméry & Gayraud, 2020) shows qualitatively that Post-traumatic stress disorder (PTSD) patients present modified language characteristics on

multiple levels. This study proposes to build a link between the language disorders presented in (Andreasen, 1979) and Traumatic language produced by PTSD subjects.

Thought disorders in speech, or more commonly speech disorders, are psychological effects that affect speech in subjects, changing the way they express themselves in a variety of ways. Since their introduction (Andreasen, 1979) they have been used to describe symptoms of mental disorders, including PTSD. Four of the eighteen speech disorders are usable on transcribed, undirected speech: illogicality, incoherence, derailment, and perseveration. (1) Perseveration: “Persistent repetition of words, ideas, or subjects so that, once a patient begins a particular subject or uses a particular word, he continually returns to it in the process of speaking.” (2) Derailment: “A pattern of spontaneous speech in which the ideas slip off the track onto another one that is clearly but obliquely related, or onto one that is completely unrelated. Things that lack a meaningful relationship may be said in juxtaposition, or the patient may shift idiosyncratically from one frame of reference to another. Sometimes, there may be a vague connection between the ideas; none will be apparent at others. This pattern of speech is often characterized as sounding "disjointed." Perhaps the commonest manifestation of this disorder is a slow, steady slippage, with no single derailment being particularly severe so that the speaker gets farther and farther off the track with each derailment without showing any awareness that his reply no longer has any connection with the question that was asked.” (3) Incoherence: “A pattern of speech that is essentially incomprehensible at times. The incoherence is due to several different mechanisms, which may sometimes all co-occur. Sometimes the rules of grammar and syntax are ignored, and a series of words or phrases seem to be joined arbitrarily and randomly. Sometimes portions of coherent sentences may be observed in the midst of a sentence that is incoherent as a whole. Sometimes the disturbance appears to be at a semantic level, so that words are substituted in a phrase or sentence so that the meaning seems to be distorted or destroyed; the word choice may seem random or may appear to have some oblique connection with the context.” Finally, (4) Illogicality: “A pattern of speech in which conclusions are reached that do not follow logically. This may take the form of non-sequiturs (ie, it does not follow), in which the patient makes a logical inference between two clauses that are unwarranted or illogical. It may take the form of faulty inductive inferences.”(Andreasen, 1979).

We aim to give statistical evidence and rich interpretation to corroborate the existence of these markers in traumatic language, while the previous study (Quillivic et al., 2023) conducted on our dataset focuses on lexical, semantic markers and speech disfluencies, we chose to conduct a deeper analysis on speech disorganization and particularly focuses on Derailment, Incoherence, Illogicality, and Perseveration.

Using precedent in text analysis by graph representation for psychiatric diagnosis (Mota et al., 2017), it appears that speech impairment effects can be shown using coherence modeling. Several other studies (Barzilay & McKeown, 2001; Grosz et al., 1995; Mesgar & Strube, 2015) used graph theory and entity relations to assess the readability of text. Our method expands on these ideas by further reducing the effects of entities in coherence assessment, leaving only the contextual coherence between sentences. We have several working hypotheses, first, based on the analysis of the literature, we posit (H1) PTSD affects speech organization and (H1.1) Criterion D is the symptom that is the cause of most speech disorders. The last hypothesis (H2), is that ruptures in the thought flow are also noticeable in the similarity distributions, which implies that Sentence Similarity Graphs can capture speech disorganization.

In this contribution, we have two main objectives. First, (O1) measures the predictive power of features derived from sentence similarity graphs (SSG) for PTSD symptoms prediction, particularly

criterion D. Second, (O2) builds an extraction method of verbatim that will help clinicians to identify speech disorders in text. We will use the testimonies from a cohort exposed to the 13th November attacks. Figure 1 presents our approach synthetically. Supplementary Information (SI) is available in `supplementary_material.docx` see Table S1 for implementation details.

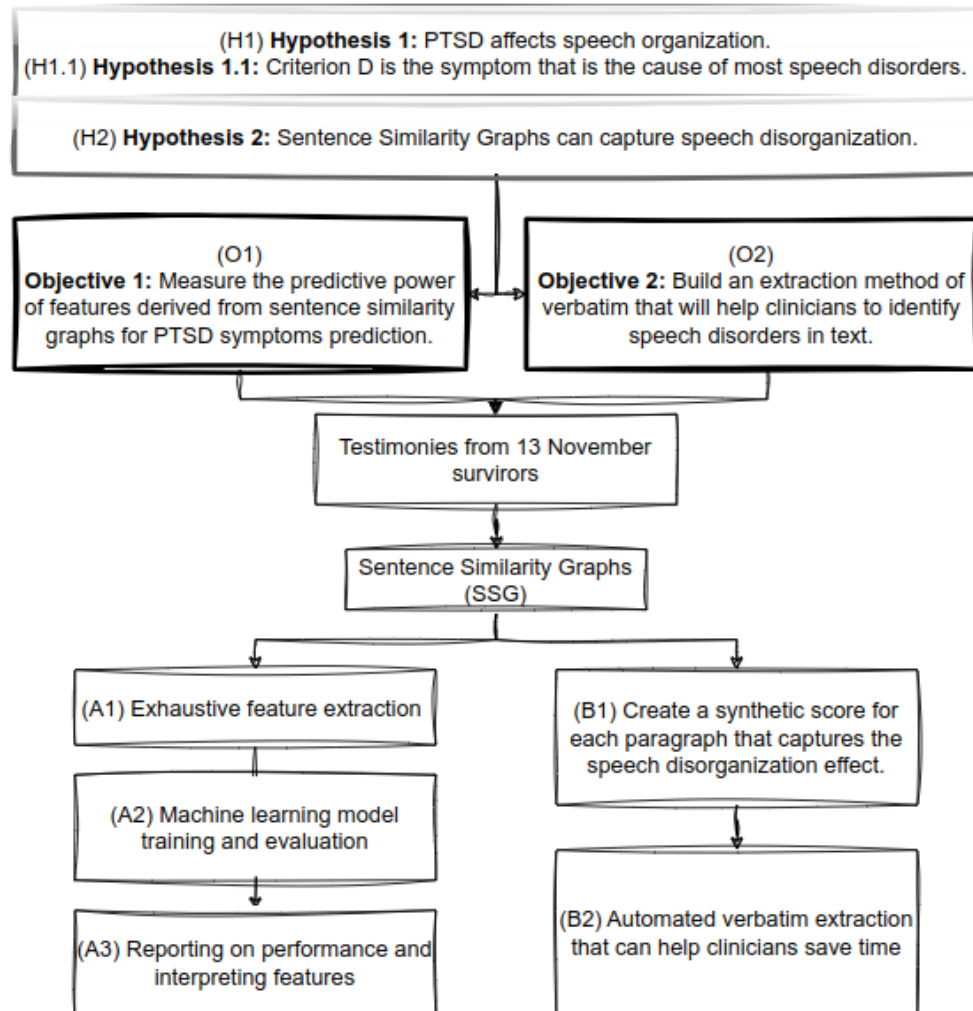


Figure 1. Summary of the approach presented in this study

## 2. Data and methods

### 2.1. Data

Our data is based on the Etude 1000, a French study that aimed to collect interviews with about 1000 people affected by the terrorist attacks in Paris on November 13, 2015. The interviews were conducted in 2016. Each interview consists of two parts, an open-ended part in which subjects are asked about their recollection of the events of November 13, and also about the cause and consequences of the event. A second set of data was constructed by asking participants to complete an emotional memory questionnaire written by psychiatrists to reveal the DSM-5 criteria for PTSD. (Eustache & Peschanski, 2022). Specifically, the data we have access to are based on 153 participants who were witnesses or victims of the terrorist attacks. Thus, we have access to 153 texts from the open-ended interview and the associated PTSD

diagnosis, with scores broken down by criterion, demographics, and exposure (victim or witness and location among the sites of the attacks). Data from the interviews were automatically transcribed and then manually corrected. Because of our sentence embedding model, we preserve the sentence structure extracted from the transcribed speech: sentences are indicated by punctuation in the original text.

## 2.2. Similarity and Model

We use a state-of-the-art sentence embedding model to compute semantic textual similarity between sentences, using this as the weights of edges in our graphs. The node of each graph is a sentence. The similarity between every pair of sentences is calculated, the graph form of this matrix is referred to as the fully connected graph. Our model of choice is the paraphrase-multilingual-MiniLM-L12-v2 (Reimers & Gurevych, 2019), which has shown top accuracy in Sentence Textual Similarity (STS) tasks among multilingual models, as well as being lighter and faster to embed than other models (Muennighoff et al., 2023) This model embeds in 384-dimensional space, which we then use to compute cosine similarity between all sentences two-by-two.

	Mean with a dot separator	Total
Words	12600 per text 26.75 per sentence	1 927 837
Sentences	533.84 per text	81678

Table 1. Statistics on the number of sentences and words in the corpus

## 2.3. Features and Graphs

To compute features, we apply two separate reconstruction methods to our graphs, to obtain partially connected graphs. A representation of a fully connected and reconstructed graph is presented in SM Figure S1. We reconstruct graphs based on a range and threshold approach, where we compute similarity up to the  $n^{\text{th}}$  sentence after the current one, and register an edge only if the similarity is over a threshold. This way, we reduce computational load and emulate the psychiatrist reading process. Each of the ranges and thresholds can have multiple values, which we limit to unbiased values based on the distribution of the dataset; in our study, the mean is 0.45, and the standard deviation is 0.1, see Figure 2. We base our analysis by comparing each text’s graph’s structure across all reconstructed values.

## 2.4. Statistical Analysis

Ensuing the last pipeline of graph reconstruction, we extract **125 usable features** and run a univariate analysis on each of our labels, to see which features are the most statistically associated with our PTSD diagnosis. From these features, we select the five most associated features for each classification problem. we average the results over 100 runs and 6 different models, to have the best possible classifier and ensure the stability of the results (see figure S2 in SM<sup>1</sup>). While all models are tested on each PTSD criterion, the shown results correspond only

<sup>1</sup>[https://docs.google.com/document/d/12jWbBVCPRn1AuNDGxMJ\\_OfQRtRPtReco/edit?usp=sharing&oid=101066242312577142518&rtfpof=true&sd=true](https://docs.google.com/document/d/12jWbBVCPRn1AuNDGxMJ_OfQRtRPtReco/edit?usp=sharing&oid=101066242312577142518&rtfpof=true&sd=true)

to the best mean score among 100 runs for the Logistic Regression model or Random Forest that are more interpretable and more used in social sciences. Figure 2 shows a summary of the Feature Extraction analysis pipeline. The coefficients of importance are all taken as the mean of all importance coefficients over 100 runs, on the best model. These coefficients show the statistical link between the feature and the label, meaning that the label being positive is more probable when the value of the feature is high.

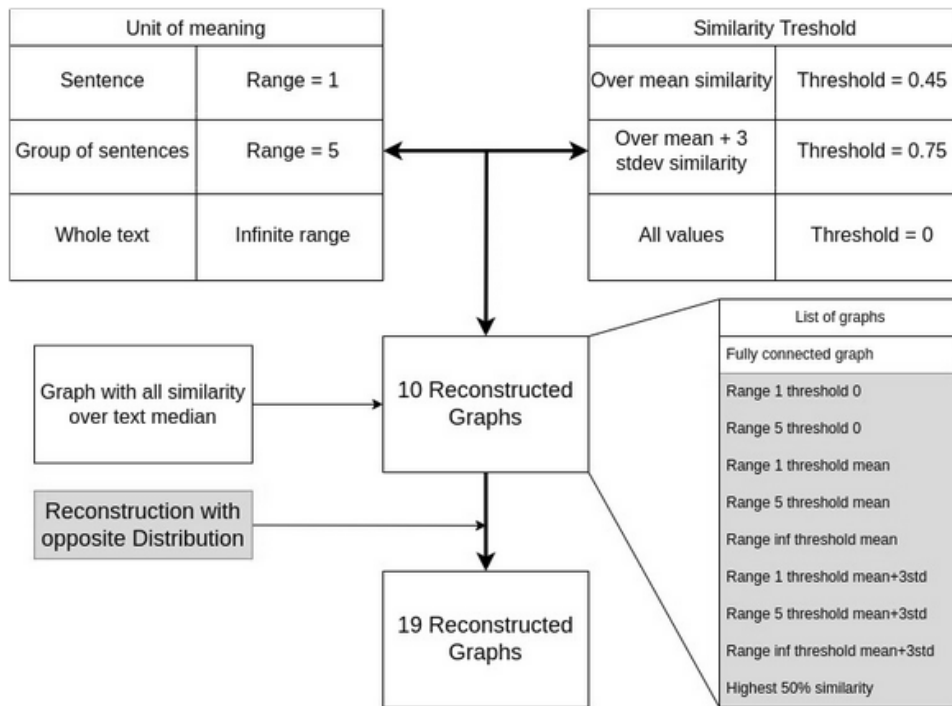


Figure 2. The figure shows the reconstruction pipeline, along with all values of both threshold and range, their explanations, and the opposite distribution graph. Using the mean and standard deviation values from the sentence similarity distribution, this process enables us to construct 19 subgraphs from which we will extract basic features.

### 2.5. Features Interpretations

From the reconstructed graphs and fully connected graphs, we computed graph-related features (see Table 2) and proposed an interpretation. In Figure 3, we propose to group some of these features and propose a possible link with language disorder (details argument in SM).

feature	Explanation	Interpretation
Density	Ratio of the number of edges on the total number of possible edges	Quantifies how dense in self-reference the narrative is
Isolates	Ratio of the number of nodes that have no edges on the total number of nodes.	Counts the number of sentences that are outliers to the described events
Density without isolates	Density computed on the graph, removing isolated nodes	Quantifies how dense the narrative is, bypassing any outliers
Number of components	Mean size of groups of nodes that are connected together	Counts the number of narrative threads in the graph
Mean similarity	Mean of similarity over graph	Gives a first estimate of how similar the narrative is sentence by sentence
Stdev of similarity	Standard deviation of similarity over graph	Gives an estimate of the normality of the distribution
Mean-Median distance	Distance between mean and median values of similarity over graph	Give an estimate of the symmetry of the distribution
Edges between mean $\pm$ 3std	Ratio of edge whose weights are between mean+3std and mean-3std	Counts the number of outliers of similarity and gives a estimate of the normality of the distribution
Edges over mean+3std	Ratio of edge whose weights are over mean+3std	Gives information about high similarity edges in outliers
Edges over mean+2std	Ratio of edge whose weights are under mean+2std	Gives information about high similarity edges

Table 2. Name, Feature Explanation, and interpretation, features that are applicable only on reconstructed graphs are highlighted in grey.

feature	Explanation	Interpretation	Category	Interpretation	Possible link with language disorder
Density	Ratio of the number of edges on the total number of possible edges	Quantifies how dense in self-reference the narrative is	Connectivity	How connected the narrative is, how rich in similarity each unit of meaning is in the text	Derailment
Isolates	Ratio of the number of nodes that have no edges on the total number of nodes.	Counts the number of sentences that are outliers to the described events			
Density without isolates	Density computed on the graph, removing isolated nodes	Quantifies how dense the narrative is, bypassing any outliers	Gaussian Normality	How Normal the distribution of similarity is, showing whether there is a equal amount of low and high similarity sentences	Illogicality, Incoherence
Number of components	Mean size of groups of nodes that are connected together	Counts the number of narrative threads in the graph			
Mean similarity	Mean of similarity over graph	Gives a first estimate of how similar the narrative is sentence by sentence	Outliers	Amount of links that are outliers to the text-specific distribution, aggregates low-similarity and high-similarity effects	Perseveration, Derailment, Incoherence
Stdev of similarity	Standard deviation of similarity over graph	Gives an estimate of the normality of the distribution			
Mean-Median distance	Distance between mean and median values of similarity over graph	Give an estimate of the symmetry of the distribution	High Similarity	Estimates the amount of high-similarity links	Perseveration
Edges between mean $\pm$ 3std	Ratio of edge whose weights are between mean+3std and mean-3std	Counts the number of outliers of similarity and gives a estimate of the normality of the distribution			
Edges over mean+3std	Ratio of edge whose weights are over mean+3std	Gives information about high similarity edges in outliers	Low similarity	Estimates the amount of low-similarity links	Derailment, Incoherence
Edges over mean+2std	Ratio of edge whose weights are under mean+2std	Gives information about high similarity edges			

Figure 3. Comparative table between feature type, category, and language disorders. In grey are features only applicable to reconstructed graphs (higher picture resolution in SM, figure S3)

## 2.6. Verbatims Extractions

To highlight parts of texts that most show the speech disorder effects. We also use the text graph and sentence similarity to extract verbatim about specific speech disorders in the text. As our similarity measure produces a score between sentences, we have to find a method to aggregate these measures in a way that gives sentences a speech disorder value. To do that, we

limit our scope to 5 sentences before and after the sentence to analyze, creating a paragraph of size 11. This paragraph induces a subgraph in our general text graph, on which we have all the similarity values. Our goal is to estimate the effect of each sentence on the local coherence. For this, we compute the difference between the mean similarity of the graph with, and without the sentence. Figure 4 summarises the calculation of the score for each sentence, taking a reduced paragraph size of 5 (2 sentences on either side) for clarity.

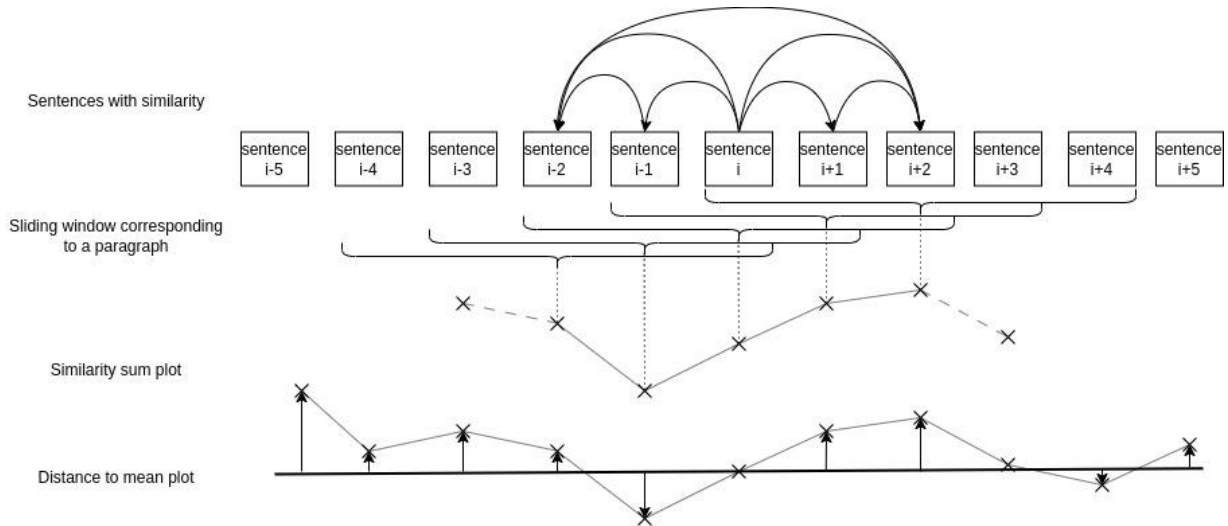


Figure 4. Summary of the Verbatim score computation method (paragraph size = 2)

### 3. Results

#### 3.1. PTSD

Over 100 runs, we achieved an AUC score of  $0.7 \pm 0.08$  which is comparable to the  $0.69 \pm 0.09$  presented in (Quillivic et al., 2023). Using the groupement of features we define, our results show that the PTSD criterion is best defined by effects related to outliers and normality (Figure 5). Because PTSD is a composite pathology with multiple criteria and symptoms, it is expected that PTSD is better defined by general effects. Indeed, the normality and outliers categories cover all speech disorders we considered. Also, we have explained (Figure 3) that outliers cover both high- and low-similarity effects, leading to the fact that PTSD is best classified by general features. The graph importance plot shows that PTSD is best revealed by graphs that compute scores within a unit of meaning (here, a paragraph), which points to the idea that our study shows effects contained in the thought structure and at a semantic level.

#### 3.2. Criterion D

Over 100 runs, we achieved an AUC score of  $0.72 \pm 0.08$  which is better than the  $0.67 \pm 0.1$  presented in (Quillivic et al. 2023). The importance of High similarity links is only confirmed by the secondary category of outliers, and the fact that all important graph reconstructions have a threshold (Figure 6). Criterion D encompasses the effects of dissociation and negative alterations in cognition and mood. High similarity and outliers encompass the effects of Perseveration, derailment, and incoherence. According to (Auxéméry et Gayraud, 2020) dissociation associated with the difficulty of going through the emotions of trauma creates narratives that always stick to the same events or words. Specifically, subjects that have a hard time reliving the events produce incoherent speech, a symbol of their trauma. Furthermore, the



graph importance plot shows that the importance is completely shared among graphs that have an infinite range, meaning that the computations of similarity are done over the whole text. This confirms that Criterion D is best described by effects that exceed the scope of a paragraph.

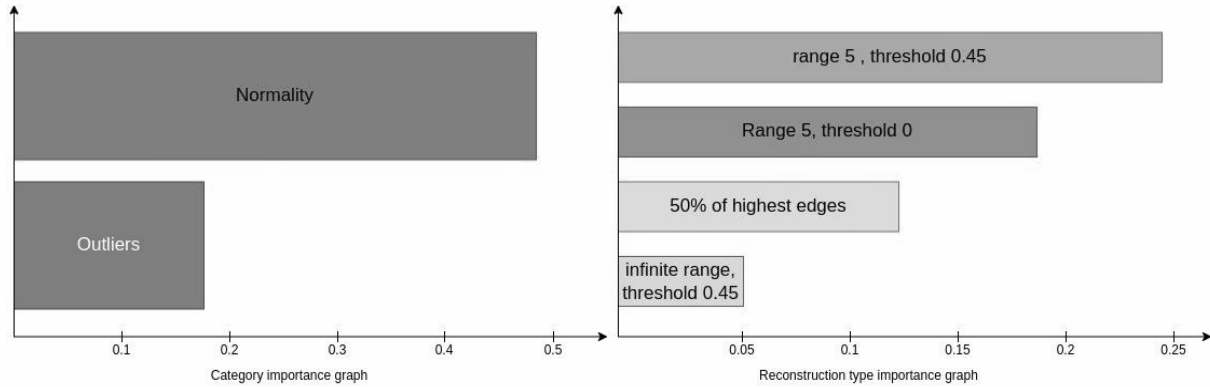


Figure 5. Category and graph importance plot for PTSD classification

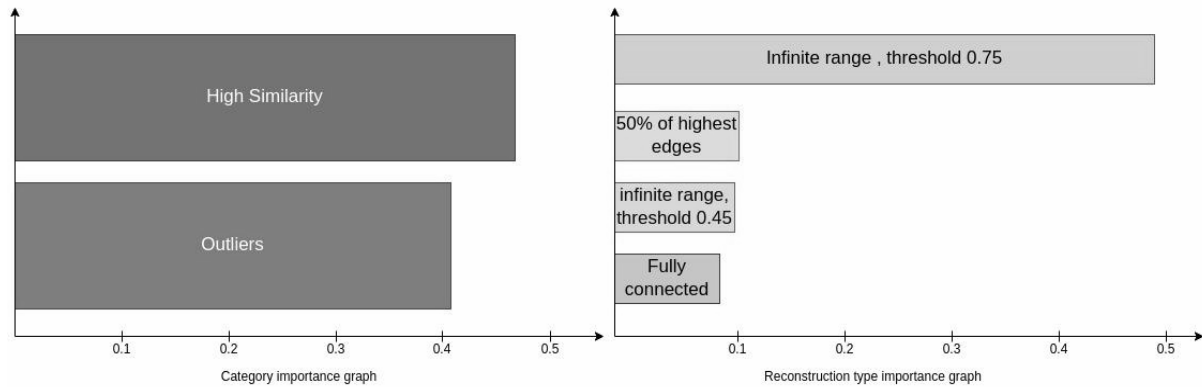


Figure 6. Category and graph importance plot for criterion D

### 3.3. Verbatims examples

Here is an example that was extracted using the methodology presented in 2.6 that is “positive”. We posit that these positive examples show **perseveration** in the text. In supplementary materials ([supplementary material.docx](#)), we give more examples showing Perseveration in text. We also extracted Negative examples showing Derailment, Illogicality, or Incoherence. Finally, as a means of comparison, we also add an example that our tool flagged as being very close to mean coherence, meaning that these examples contain no speech disorder.

Positive Verbatim extracted from interview of subject PAR0480 of scaled score + 5.20 :

*Je dis: "Je sais pas" Je dis: "Quand je suis sortie il y en avait pas beaucoup des flics hein". J'ai dit : "Je pense que c'est peut-être les gars qui j'ai l'impression qu'il posaient leur arme ou des armes comme si quelqu'un posait les armes qu'ils faisaient du scotch je sais pas ce que c'était enfin". Et je dis : "Il y a des il y a un moment il m'a il a marché derrière moi donc". Je dis: "'J'ai eu peut-être la chance" je lui expliquais. Le gars il marchait derrière moi peut-être sur la rambarde moi en quand je suis tombée j'étais vraiment pas loin de la rambarde du coup le gars quand il tirait il tirait au loin comme ça et il a pas tiré par-dessus la rambarde pour tirer comme ça. Donc j'ai dit : "Du coup j'ai eu du bol parce que je veux dire". Enfin je lui ai expliqué avec les positions et j'ai dit: "Bon voilà quand ils sont arrivés à l'entrée ils ont arrosé comme ça. Donc même quand je suis partie j'ai vu les tas de cadavres tas de cadavres là tas de cadavres là". A un moment je dis: "Bah c'est j'étais là. Ouais bah tous ceux que j'ai vus ben ouais j'étais à peu près dans ce sens -là j'ai eu peut-être la chance de tomber de ce côté -là".*

#### 4. Conclusion and Discussion

The present study represents a pioneering attempt to establish a relationship between language disorders, as described by Andreasen (1979), and PTSD, with a specific focus on Criterion D symptoms. Although the study goals outlined in the introduction were only partially met, significant progress was made. First, we successfully developed features with substantial predictive power, resulting in a respectable classification AUC score. Specifically, we improved classification performance on criterion D, which includes negative changes in cognition and mood, and dissociation, using only features derived from the Sentence Similarity Graph (SSG). This result supports the hypothesis (H1.1) suggesting that criterion D predominantly affects speech organization. Analysis of the classifiers provided valuable insights into the specific links influenced by criterion D within the graph, highlighting in particular the link between dissociation, mood change, and perseveration. Furthermore, we propose a correlation between SSG features and four different speech disorders (illogicality, incoherence, derailment, and perseveration). Although we provide a detailed argument explaining the construction of this correlation in the Supplementary Material (SM), further refinement is needed. Future efforts should focus on identifying traits that more closely match the definitions of the disorders, while also emphasizing conceptual simplicity. The delineation between derailment, incoherence, and illogicality requires finer differentiation, which could be facilitated by the use of generative technology to produce synthetic datasets for quantitative validation of the alignment of the proposed features with their respective definitions.

Concerning the second objective (O2), the proposed scoring mechanism provides a straightforward means of distinguishing perseveration from illogicality, incoherence, and derailment. This score holds promise as a basic tool for clinical verbatim extraction, allowing for accelerated analysis without compromising the integrity of the patient's discourse or the clinician's judgment. However, the validity of this scoring system warrants further scrutiny, with future research requiring a methodological approach that promotes generalizability across diverse datasets. Despite the careful application of machine learning methods, the possibility of overlearning corpus characteristics due to sample size cannot be ruled out.

Notably, the proposed methodology is pathology and language agnostic, as it does not rely on external lexical resources. The feature-building approach centered on perseveration,

incoherence, and derailment, based on SSG, holds promise for applicability in diverse contexts beyond the scope of this study.

## References

- American Psychiatric Association (s. d.). *APA - The Structured Clinical Interview for DSM-5®*.  
<https://www.appi.org/products/structured-clinical-interview-for-dsm-5-scid-5>
- Andreasen N.C. (1979). Thought, language, and communication disorders. I. Clinical assessment, definition of terms, and evaluation of their reliability. *Archives of General Psychiatry*, 36 (12), 1315-1321.  
<https://doi.org/10.1001/archpsyc.1979.01780120045006>
- Barzilay R. and McKeown K.R. (2001). Extracting paraphrases from a parallel corpus. In *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics - ACL '01*, Toulouse, 50-57. <https://doi.org/10.3115/1073012.1073020>
- Eustache F. and Peschanski D. (2022). Toward new memory sciences : The Programme 13-Novembre. *Prog Brain Res*, 274 (1), 177.  
<https://doi.org/10.1016/bs.pbr.2022.07.003>
- Grosz B.J., Joshi A.K. and Weinstein S. (1995). Centering : A Framework for Modeling the Local Coherence of Discourse. *Computational Linguistics*, 21(2), 203-225.
- Mesgar M. and Strube M. (2015). Graph-based Coherence Modeling For Assessing Readability. In Palmer M., Boleda G. and Rosso P. (Eds.), *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics*, Denver, 309-318. <https://doi.org/10.18653/v1/S15-1036>
- Mota N.B., Copelli M. and Ribeiro S. (2017). Graph Theory Applied to Speech : Insights on Cognitive Deficit Diagnosis and Dream Research. In Poibeau T. and Villavicencio A. (Eds.), *Language, Cognition, and Computational Models*, Cambridge University Press, 81-98.  
<https://doi.org/10.1017/9781316676974.004>
- Muennighoff N., Tazi N., Magne L. and Reimers N. (2023). MTEB : Massive Text Embedding Benchmark. *arXiv:2210.07316*. <http://arxiv.org/abs/2210.07316>
- Quillivic R., Gayraud F., Auxéméry Y., Vanni L., Peschanski D., Eustache F., Dayan J. and Mesmoudi S. (2023). Towards Unlocking the Linguistic Code of Post-Traumatic Stress Disorder : A Comprehensive Analysis and Diagnostic Approach. <https://doi.org/10.21203/rs.3.rs-3747882/v1>
- Reimers N. and Gurevych I. (2019). Sentence-BERT : Sentence Embeddings using Siamese BERT-Networks. *arXiv:1908.10084*. <https://doi.org/10.48550/arXiv.1908.10084>

## Fundings

The study was conducted within the framework of the “Programme 13-Novembre”, funded by the French General Secretariat for Investment (SGPI) through the National Research Agency (ANR) and the “Programme d’investissement pour l’Avenir” (PIA ANR-10-EQPX-0021-01). It also receives the support of Region île-de-France through deliberation n°CP2021-297.