

# OBSER-MENH: Digital OBSERvatory of MENTal Health in Social Networks for Healthcare Institutions Based on Language Technologies\*

Alicia Pérez<sup>1,\*</sup>, Maite Oronoz<sup>1</sup>, Juan Martinez-Romo<sup>2,3</sup> and Lourdes Araujo<sup>2,3</sup>

<sup>1</sup>HiTZ Basque Center for Language Technologies - Ixa (UPV/EHU), Manuel Lardizabal 1, 20018 Donostia, España (<http://www.hitz.eus>)

<sup>2</sup>NLP & IR group (UNED). C/ Juan del Rosal, 16, 28040 Madrid, España (<http://nlp.uned.es/>)

<sup>3</sup>Instituto Mixto UNED-ISCHII (IMIENS)

## Abstract

This project highlights the importance of analyzing social networks in the digital transition that must take place in the administration in order to adapt to the new sources of information. The creation of the mental health observatory proposed could help hospitals to: receive a global image of the social perception and its changing dynamics as a source of social information, manage their human resources by anticipating work peaks based on the analysis of trends found in social networks, consult demographic profiles, obtain an analysis of temporal or seasonal characteristics, and discover relationships. This project makes technological contributions in: the generation and availability to the community of i) linguistic resources; and ii) tools to carry out natural language understanding adapted to Spanish and the terminology of these networks. The main challenge of the project is to convert unstructured textual information into interpretable knowledge and story-telling.

## Keywords

Mental health, social networks, health care institutions, natural language processing

## 1. Introduction

According to the World Health Organization, 80 people die by suicide every hour, being suicide among the 10 most common causes of death globally. In Spain, in 2022, 2015 people died due to suicide and self-harm injuries while 761 were victims of traffic accidents, that is, last year the number of suicides almost tripled that of victims of traffic accidents [1]. The correct identification of signals that could indicate suicide risk is a core, yet undisclosed, cornerstone. Some risks and factors connected to suicide ideation or commitment include: mental health disorders, loneliness or lack of connection with significant ones, depression, bullying, consumption of substances, lack of adherence to treatment, self-harm etc. There are documented differences in the population not only by health condition but also by socio-economic factors [2], job occupation, age (young [3] and elder people [4]), gender and other social and demographic factors.

In the foreword of the “Global Governance Toolkit for

Digital Mental Health” document launched by World Economic Forum and Deloitte [5], it is stated that “Disruptive technology in mental health provides an opportunity to create breakthrough solutions that improve mental health and well-being outcomes on a greater scale than ever before”. It is not new the fact that artificial intelligence leverages valuable knowledge from collective behaviors as it is the case of the social observatories in politics [6].

Our hypothesis rests on the fact that collective emotional and behavioral reflections in the social media entail inherent psycho-linguistic features connected to general well-being. In our case, we will pay special attention to risk factors such as self-harm, loneliness, depression and anxiety. Moreover, changes in these features might reveal changes that can light pathways to aid healthcare practitioners envisaging a punctual social picture related to mental illnesses. The work carried out by Fine et al. [7] and [8] is a proof of concept of how publicly available social media data might be used to assess population-level mental health. The analysis of data outside the healthcare system, and in particular social media data, may lead to the development of automatic systems for assessing suicide risk based on what people share on their social networks. This kind of information allows looking into a person’s life in a more frequent and daily basis, when the potential risk of committing suicide can be monitored much more efficiently [9]. We find, however, a gap for this type of approaches involving deep natural language understanding in Spanish.

SEPLN-PD 2023: Annual Conference of the Spanish Association for Natural Language Processing 2023: Projects and System Demonstrations

\*Corresponding author.

✉ alicia.perez@ehu.eus (A. Pérez); maite.oronoz@ehu.eus

(M. Oronoz); juaner@lsi.uned.es (J. Martinez-Romo);

lurdes@lsi.uned.es (L. Araujo)

📞 0000-0003-2638-9598 (A. Pérez); 0000-0001-9097-6047

(M. Oronoz); 0000-0002-6905-7051 (J. Martinez-Romo);

0000-0002-7657-4794 (L. Araujo)

© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License

Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

OBSER-MENH project<sup>1</sup> involves advanced computational semantic technologies and makes a technological contribution generating tools to carry out natural language understanding adapted to Spanish and the terminology of these networks with the added value of rendering unstructured textual information into interpretable knowledge.

## 2. Objectives

In brief, a review to the antecedents outlined the importance of detecting negation, pronouns, variations in the emotional state, abstract adjectives, social factors etc. Thus, tools should be adapted to Spanish in this domain. All in all, we find that there is a gap in Spanish psico-linguistic analysis. We miss annotated data and also the application of medical entity recognition (MER) prior to polarity discovery approaches. Having detected the strengths and weaknesses of the related work, we focus on the gaps, and thus define the objectives. General objectives include:

- Development of tools for information retrieval in Spanish focused on the analysis of data from social networks related to mental health for the early detection of changing trends in psycho-linguistic characteristics.
- Analysis of existing public data on mental health, development of annotation schemes for the problems addressed in the sub-projects (including gender data).
- Creation and adaptation of NLP tools to the domain of mental health in social media.
- Generation of demographic profiles focusing on the psycho-linguistic characteristics expressed in the texts and inherent to potential mental health risks for the benefit of individuals, taking gender dimension into account.
- Analysis of the relationships among the different problems addressed in the sub-projects that, through artificial intelligence, provide dynamism for the benefit of better healthcare.

## 3. Novelty and added value

A careful review to antecedents disclosed that eHealth in social media has been addressed in English. Our project will bridge the gap to adapt deep natural language understanding tools to Spanish and has an added value to alternative Spanish variants (e.g. Argentina, Chile, Mexico etc.) with the particular analysis of each variant by means of the demographic-profile analysis. Social media

data is the core of our project on which rely the proposed artificial intelligence approaches. Annotating data aided by human experts provides extra possibilities for the algorithms to learn in semi-supervised fashions or following few-shot learning schemes. However, general purpose tools to process natural language are still to be adapted to the nuances of language employed in social media. A contribution of this project rests on the re-training or adaptation of tools to analyse and process social-media texts. Specific lexica will be created to suit domain adapted language models that are employed by deep neural language understanding approaches such as BERT and related transformers. These tools are crucial to gain insights beyond static linguistic lists and represent a step ahead in semantic relatedness. Indeed, it is a deep language understanding the approach fostering the novelty of our approach, an observatory displaying psycho-linguistic features and revealing chronological dynamics in the embedded topics, discovering latent demographic profiles of anxiety, loneliness, depression, self-harm risk and suicide ideation. The novelty rests, not only on the information extracted from unstructured media but also in the purpose itself aiming at providing insights from language to experts in Mental eHealth and bridging the gap between users and specialised health facilities and associations. In fact, the project shall contribute with suited negation and speculation detection tools, medical entity recognition and emotional language discovery. The valuable resources developed in this project shall be reported and left to the research community.

In our project, the added value in terms of return to the society rests on the formulation of the time component within an observatory in order to keep track and, particularly, focus on changes across time in social perception, emotions and behaviour latent in language and building demographic profiles of latent linguistic features. Moreover, we envisage to render unstructured linguistic data into human-friendly graphical data to aid decision support by health practitioners.

The expected findings would be in the research line of computational social psychology, particularly, in psycholinguistics. In brief, user generated content is a valuable means of collective intelligence and subjective perception of reality, it is dynamic and fast. Connective action and post sharing leads to under-used complex though valuable big data in terms of natural language.

## 4. Methodological approach

The methodology proposed to cope with OBSER-MENH is broadly depicted in Figure 1. In brief, in order to cope with the aforementioned goals, the design encompasses the following modules and tasks:

- **Data:** this module involves two important tasks:

<sup>1</sup>OBSER-MENH project: <http://nlp.uned.es/obser-menh-project>

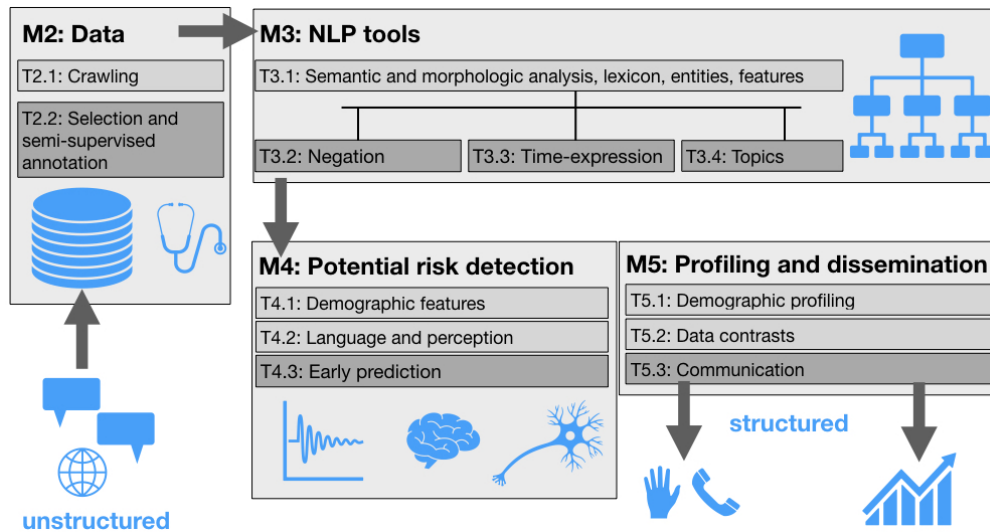


Figure 1: OBSER-MENH: methodological overview

1. Crawling a large data-set from social networks (e.g. Twitter) in Spanish.
  2. Health-expert aided manual and semi-automatic annotation of relevant linguistic cues e.g. drugs, key phrases, specific lexicons (e.g. chapter F of ICD-10 or, alternatively, Snomed-CT) and gender indicators.
- **Adaptation of general NLP tools:** such as morpho-linguistic and semantic analysis; negation detection; polarity; gender identification; implementation of embedded lexica; embedded topic discovery; transfer-based deep neural natural language modeling and understanding approaches; assessment with health-experts and linguists.
  - **Potential risk detection:** Implementation of the response from the observatory with the following pillars: 1) early prediction of suicide ideation, self-harm risk, perception of severe loneliness, anxiety etc. and, accordingly, provide, in banners, contrasted means of helpful resources [10] to the users (health systems and expert associations); 2) graphic representation of topical evolution across time to model trends and communicate the observations to health professionals.
  - **Profiling and dissemination:** generation of demographic and psycho-linguistic profiles related to the analyzed problems together with expert-aided evaluation of the generated profiles and the extracted patterns on the considered mental problems.

In brief, given unstructured data from social media, the aim is to attain structured information in terms of

changing emotions and behaviors potentially connected with substantial changes in mental health conditions.

## 5. Research groups

The OBSER-MENH project is arranged as two coordinated sub-projects (each led by a research group described latter) entitled as follows:

- “GELP: Generation with Language Processing of demographic profiles in social networks for detection of suicide risk and its relationship with other psychological problems”. The project focuses on the potential risk of suicide opening a way to self-harm detection and other conditions related to suicide or suicide ideation.
- “LOTU: Analysis of psycho-linguistic features for early detection of changes in social media about Loneliness and isolation perception employing deep natural language understanding”. The main focus of this sub-project is the detection of psycho-linguistic elements reflecting situations of loneliness and isolation and suicidality-related issues such as depression or anxiety.

OBSER-MENH is a multi-disciplinary project and, accordingly, it is supported by research groups and experts in different fields, with the flagship being Natural Language Processing applied to the Clinical domain together with clinicians and experts in the field of psychiatry within the Public Health System as described in what follows:

- **NLP&IR group at UNED** (<http://nlp.uned.es/>) Has a long trajectory in Intelligent Access to Information and Knowledge Representation. The group counts on several research lines within the clinical domain and demographic cue extraction from text (of particular relevance in the segmentation study underlying mental health). NLP&IR group at UNED is the leader of the aforementioned GELP sub-project.
- The psychiatry service of the **Hospital Clínico San Carlos** in Madrid, cooperates with NLP&IR group at UNED in the GELP project.
- **HiTZ** (<http://www.hitz.eus>) HiTZ is a reference center on Language Technologies aiming at the promotion of research, training, technological transfer and innovation in Artificial Intelligence focused on language and speech technologies. The subset of researchers from HiTZ involved in this project are all within Ixa research group (<http://www.ixaeus.com>) a consolidated group with beyond than 35 years of existence in which a research line started more than 12 years ago is devoted to NLP in the clinical domain. HiTZ is the leader of LOTU sub-project.
- **Osakidetza**, the Public Health System within the Basque Country and, specifically, experts within the Network of Mental Health in Biscay involved in BioCruces (the Health Research Center of Biscay) cooperate with HiTZ in the LOTU project.

## 6. Conclusions

The creation of the mental health observatory presented in this project could help to: (1) receive a global image of the social perception and its changing dynamics as a source of social information, (2) manage their human resources by anticipating work peaks based on the analysis of trends found in social networks, (3) consult demographic profiles, (4) obtain an analysis of temporal or seasonal characteristics, and (5) infer latent relationships between different problems.

## 7. Acknowledgements

OBSER-MENH, with subprojects GELP (TED2021-130398B-C21) and LOTU (TED2021-130398B-C22) are funded by MCIN/AEI/10.13039/501100011033 and by the European Union "NextGenerationEU"/PRTR. In addition, this work was partially funded by the Spanish Ministry of Science and Innovation (DOTT-HEALTH/PAT-MED PID2019-106942RB-C31 and INDICA-MED PID2019-106942RB-C32) by the Basque Government (IXA IT-1570-22); and by EXTEPA within Misiones Euskampus 2.0.

## References

- [1] Instituto Nacional de Estadística, Defunciones por causa de muerte, <https://www.ine.es/jaxi/Tabla.htm?tpx=55863&L=0>, 2022. Accessed: 2023-03-06.
- [2] F. Ferretti, A. Coluccia, Socio-economic factors and suicide rates in European Union countries, *Legal Medicine* 11 (2009) S92–S94.
- [3] D. Wasserman, Q. Cheng, G.-X. Jiang, Global suicide rates among young people aged 15-19, *World psychiatry* 4 (2005) 114.
- [4] M. Waern, E. Rubenowitz, B. Runeson, I. Skoog, K. Wilhelmson, P. Allebeck, Burden of illness and suicide in elderly people: case-control study, *Bmj* 324 (2002) 1355.
- [5] World Economic Forum, Deloitte, Global governance toolkit for digital mental health: Building trust in disruptive technology for mental health, 2021. URL: [https://www3.weforum.org/docs/WEF\\_Global\\_Governance\\_Toolkit\\_for\\_Digital\\_Mental\\_Health\\_2021.pdf](https://www3.weforum.org/docs/WEF_Global_Governance_Toolkit_for_Digital_Mental_Health_2021.pdf).
- [6] S. Caton, M. Hall, C. Weinhardt, How do politicians use Facebook? An applied social observatory, *Big Data & Society* 2 (2015) 2053951715612822.
- [7] A. Fine, P. Crutchley, J. Blase, J. Carroll, G. Coppersmith, Assessing population-level symptoms of anxiety, depression, and suicide risk in real time using NLP applied to social media data, in: *Proceedings of the fourth workshop on natural language processing and computational social science, 2020*, pp. 50–54.
- [8] H. Moradian, M. A. Lau, A. Miki, E. D. Klonsky, A. L. Chapman, Identifying suicide ideation in mental health application posts: A random forest algorithm, *Death Studies* (2022) 1–9.
- [9] P. Resnik, A. Foreman, M. Kuchuk, K. Musacchio Schafer, B. Pinkham, Naturally occurring language as a source of evidence in suicide prevention, *Suicide and Life-Threatening Behavior* 51 (2021) 88–96.
- [10] M. K. Nock, E. M. Kleiman, M. Abraham, K. H. Bentley, D. A. Brent, R. J. Buonopane, F. Castro-Ramirez, C. B. Cha, W. Dempsey, J. Draper, et al., Consensus statement on ethical & safety practices for conducting digital monitoring studies with people at risk of suicide and related behaviors, *Psychiatric research and clinical practice* 3 (2021) 57–66.