# Research statement: Computational characterization of mental states: a natural language processing approach

**PhD Student**: Facundo Carrillo, e-mail: fcarrillo@dc.uba.ar

Laboratorio de Inteligencia Artificial Aplicada, Departamento de Computación, FCEyN, UBA, Pabellón 1, Ciudad Universitaria, 1428 Buenos Aires, Argentina

The massive (and increasingly accelerating) digital availability of thought products in textual format opens a window to study the brain and mind in radically novel ways. We propose that through the use of natural language tools on text and state-of-the-art mathematical approaches we may assess mental states with unprecedented detail and precision.

While biological psychiatry and neuroscience seek to understand the brain and the mind, they face the problem of bridging the large gap between the micro- and macroscopic levels. Historically, behavioral studies have lacked the rigor and sophistication of neuroscientific research. However, the cognitive sciences field has recently developed an approach to amass enormous amounts of data through game-like web applications. Together with the access to large web repositories of text, and to practically unlimited computational power, these developments are changing the way we can characterize cognition and behavior.

We focused in mental health study, most psychiatric evaluations are presently done by qualitative analysis of interviews in hospitals or mental institutions. Recently, we have detected stereotyped markers in text that may distinguish between psychosis pathologies (mania and schizophrenia). With the development of better analytic methods for detection of these markers, we propose the generation of quantitative (and semi-automatic) techniques that provide the specialist with additional information taken from written text to help in the diagnosis and tracing of patients.

The goal of this plan is the development and use of machine-learning techniques to study massive-scale digital text corpora associated with cognitive processes, aiming at identifying the mental operations underlying behavioral processes. The goal is two fold: first, to decode the regularities in the corpora to infer generative rules of human thought, and second, to incorporate them in artificial intelligence models.

Written text is the maximum exponent of this new reality. Every day, web pages, books, exams, interviews are being produced and stored in digital format, e.g. on the Internet. Based on these repositories, analytic strategies have been investigated which recover the semantic structure of words. Also, machine-learning techniques have been developed to analyze texts and obtain hidden information about them, such as the topics involved or the mental state of the author [1].

One of the most complicated issues is the availability of methods and tools in languages other than English. In previous research we observed that this

issue may sometimes be skipped via machine translation methods, obtaining results similar to manual translation. Still, the adaptation of natural language processing tools into a general framework with texts from different repositories in many languages (e.g. Spanish and Portuguese) remains an important aspect of this research program.

Most text analysis tools rely on a training corpus, generally curated by experts. With the massive digital availability of texts in most of the world's main languages (digital newspapers, blogs, web pages) we propose the creation of a framework that may incorporate selected sources of text from the Internet as training corpus and generate a semantic space in a multi-language perspective.

With these tools, we propose the identification of new markers to contribute to the diagnosis and treatment of psychiatric patients.

During the last year we created and tested some methods to characterize mental alteration. For this, we implemented a new method to identified the speech changes produced by drug intoxication effects [2] and we tested in a real case.

At this time we are working in a new multi-language semantic method based on Twitter. This method is similar to Google Similarity Distances [3] but with more resolution time. This features allows to study how particulars events change the semantic network.

Also we are working in other medical application: Automated analysis of speech predicts transition to psychosis in high-risk patients.

## References

1. Dumais, S.T.: Latent semantic analysis. Annual review of information science and technology **38**(1) (2004) 188–230
2. Bedi, G., Cecchi, G.A., Slezak, D.F., Carrillo, F., Sigman, M., de Wit, H.: A window into the intoxicated mind? speech as an index of psychoactive drug effects. Neuropsychopharmacology (2014)
3. Cilibrasi, R.L., Vitanyi, P.M.: The google similarity distance. Knowledge and Data Engineering, IEEE Transactions on **19**(3) (2007) 370–383