

## **Can we detect preclinical Alzheimer's disease using smartphone data?**

L'UCLouvain (Belgique) recherche un chercheur en traitement automatique du langage ou profil similaire pour un projet de 2 ans minimum (possibilité de réaliser une these ou non). Engagement à partir du 1<sup>er</sup> février 2019.

Description du projet :

Over the last twenty years, the scientific community has made tremendous progress to detect Alzheimer's disease (AD) before the onset of clinical symptoms. Medical analyses require an invasive puncture or are very expensive. Thus, developing affordable and minimally invasive screening tools to detect AD pathology in the general population would be of great help for both research and clinical applications. We answer this need by proposing an automated analysis of linguistic features stored in electronic messages. Indeed, smartphones now keep 'sent messages' for several years, allowing us to analyse retrospectively longitudinal linguistic changes in narrative written data. After consent, we will download all 'sent messages' from the Preclinical AD patients' phones together with the date and time information associated with the messages. We will then use tools and methods from Natural Language Processing (NLP) to analyse this personal big-data.

The 2-year job will rely on the following work packages:

1. Lexical components: Lexicometric techniques will be used to grasp the richness of vocabulary and the use of empty words: metrics as type/token and hapax legomena ratio will be employed.
2. Syntactic and spelling components: We will statistically evaluate each patient's ability to produce or avoid spelling errors through time. It should be noted that we have developed an expertise in differentiating spelling mistakes from word play.
3. Punctuation and other components: Particular attention will be paid to the use of punctuation as we know that electronic environments stimulate a different use of these elements.
4. Modelisation: we will then have to modelise the graphs with the help of different tools, such as Unitex grammars (Paumier, 2016).
5. Machine learning algorithms will be developed to learn linguistic features through the automatic processing of productions over time. Diachronic linguistics considers language in its temporal context, according to what precedes and follows. Statistical approaches will include text classification resources as features within a support vector machine (SVM).

Profil recherché :

- Diplôme TAL ou assimilé
- Expérience ou intérêt pour les neurosciences
- Natif francophone
- Niveau anglais min. B2 (+ expérience en lecture d'articles scientifiques B2)

Posez des questions ou envoyez directement votre curriculum vitae (et la liste de vos publications, si vous en avez) dans un mail expliquant votre motivation pour ce poste à : [louise-amelie.cougnon@uclouvain.be](mailto:louise-amelie.cougnon@uclouvain.be), directrice de la recherche au Media Innovation and Intelligibility Lab, avant le 15 janvier 2018.

Le projet est issu d'une collaboration entre le MiiL (centre d'innovation médiatique), le Cental (centre de traitement automatique du langage) et les Cliniques universitaires Saint-Luc, à l'UCL.