

# JEAN-BAPTISTE TANGUY

07 66 03 43 12 ◊ jean-baptiste.tanguy@sorbonne-universite.fr

## FORMATION

---

<b>Doctorat - LabEx OBVIL, CELLF, STIH</b>	2019 - ...
Sorbonne Université, faculté des Lettres, Paris (75)	
<b>Master de Langue et Informatique</b>	2016 - 2019
Sorbonne Université, faculté des Lettres, Paris (75)	
<b>Bachelor en Sciences de l'ingénieur</b>	2015 - 2016
Ecole Nationale Supérieure de Cognitique, Bordeaux (33)	
<b>CPGE Lettres et Sciences Sociales (Khâgne B/L)</b>	2013 - 2015
Lycée Saint François-Xavier, Vannes (56)	

## PRINCIPAUX TRAVAUX EN TRAITEMENT AUTOMATIQUE DES LANGUES

---

**Stage de six mois, Credit Agricole SA, DataLab, Montrouge (92)** 2019  
Partie R&D : amélioration de *word embeddings* par *retrofitting* (utilisation de *word2vec* et *GloVe*).  
Partie applicative : évaluation et amélioration du socle linguistique du DataLab (annotation de corpus, règles linguistiques, identification automatique de concepts).

**Stage de six mois, Advanced Decision, Paris (75)** 2018  
Réalisation d'un prototype d'extraction d'opinions multi-aspects (thèmes et sentiments) en utilisant des CRF (*Conditional Random Fields*). Création des données d'apprentissage, de la chaîne de traitements et de l'ingénierie des tests.

**Mémoire, Sorbonne Université, Paris (75) Correction automatique de réponses d'étudiants à des QROC** 2017  
Synthèse de l'état de la recherche contemporaine sur ce domaine. Développement d'une méthodologie de correction automatique et d'évaluation de trois techniques (TF-IDF, grammaires hors-contexte et apprentissage automatique). Conception d'une application en Python puis en C++.

### Projets, Sorbonne Université, Paris (75)

Extraction endogène de termes, indépendante de la langue, dans un corpus trilingue (français, anglais, polonais).

Réalisation d'un moteur de recherche sur les médicaments. Utilisation de solr (Lucene).

Classification de textes par style littéraire. Utilisation de classifieurs.

Analyse langagière des discours de campagne des candidats aux élections présidentielles 2017 (TXM et Antconc).

## COMPÉTENCES

---

**Traitement Automatique des Langages** Talismane, NLTK, Spacy, Unitex, TXM, Gate, Antconc  
**Machine Learning** Weka, sklearn, wapiti

**Représentation/Recherche des connaissances** Protege, solr (Lucene)

**Informatique** Python, Java, C++, C# - XML, XSLT - HTML, CSS, PHP, SQL

**Langues** Anglais : C1 - Espagnol : C1