

JEAN-BAPTISTE TANGUY

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

FORMATION

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

SUJET DE THÈSE

Financé par la région Île-de-France (DIM Sciences du Texte et Connaissances nouvelles), et en partenariat avec la Bibliothèque Mazarine, je réalise ma thèse au sein du projet Antonomaz. Cette thèse est à mi-chemin entre le Traitement Automatique des Langues Naturelles (TALN) et les Humanités Numériques (HN). Elle est dirigée par Glenn Roe et encadrée Karine Abiven et Gaël Lejeune. Des motivations patrimoniales aux usages en HN, je m'intéresse à l'acquisition et l'exploitation des données textuelles numérisées - des imprimés « anciens ». Mon corpus de travail est celui des Mazari-nades, imprimées pendant la Fronde (1648-1652). Les questions de constitution de corpus numériques, de textualité mais aussi d'O.C.R. et de corpus bruités sont au cœur de mon travail.

EXPÉRIENCE EN T.A.L.

Stage de six mois, Credit Agricole SA, DataLab, Montrouge (92) Partie R&D : amélioration de <i>word embeddings</i> par <i>retrofitting</i> (utilisation de <i>word2vec</i> et <i>GloVe</i>). Partie applicative : évaluation et amélioration du socle linguistique du DataLab (annotation de corpus, règles linguistiques, identification automatique de concepts).	2019
Stage de six mois, Advanced Decision, Paris (75) Réalisation d'un prototype d'extraction d'opinions multi-aspects (thèmes et sentiments) en utilisant des CRF (<i>Conditional Random Fields</i>). Création des données d'apprentissage, de la chaîne de traitements et de l'ingénierie des tests.	2018
Mémoire, Sorbonne Université, Paris (75) Correction automatique de réponses d'étudiants à des QROC 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++.	2017

PUBLICATIONS

Standardizing linguistic data: method and tools for annotating (pre-orthographic) French, S. Gabay, J.-B. Tanguy, T. Clérice, J.-B. Camps - DDH, octobre, 2020, Hammamet (à paraître)

Exploiter des modèles de langue pour évaluer des sorties de logiciels d'OCR pour des documents français du XVIIe siècle, J.-B. Tanguy - RECITAL, juin, 2020, Nancy

Analyse d'opinions multi-aspects pour la recommandation fine de restaurants, I. Tellier, H. Hammouche, D. Cholvy, J.-B. Tanguy - CNIA & RJCIA 2018