

PROPOSITION SUJET DE THÈSE CONTRATS DOCTORAUX - 2023-2026

X Contrat doctoral fléché FR Agorantic
X Contrat doctoral fléché EUR InterMEDIUS

Pour candidater sur ce sujet, les personnes intéressées doivent contacter le/la directeur.rice de thèse avant le **2 juin 2023**.

Les auditions des candidats retenus auront lieu début juillet.

Directeur.rice de thèse :	Juan-Manuel Torres	Mail :	Juan-manuel.torres@univ-avignon.fr
Laboratoire :	LIA	Téléphone :	0490843568
Co-directeur.rice et/ou encadrant.e :	Graham Ranger Elsa Cavalie (encadrante)		
Laboratoire :	ICTT		

Titre en français : Génération automatique de texte : un système hybride combinant apprentissage profond, linguistique et statistique

Titre en anglais : Automatic text generation: a hybrid system combining deep learning, linguistics and statistics

Résumé en 7 lignes :

Cette thèse cherche à développer des algorithmes de production de texte ayant des caractéristiques à la fois de littérarité et de créativité dans le cadre de récits en français. Nous allons nous appuyer sur la combinaison de méthodologies TAL telles que la génération de texte en boîte (canned text et shallow parsing), recherche d'information (RI), corpus ainsi que sur la production de phrases au moyen de techniques d'apprentissage profond (IA et RNN du type transformateur).

Mots clés :

1- Présentation du sujet

Traitement Automatique des Langues ; Corpus littéraires ; Machine Learning ; Deep Learning ; Intelligence artificielle

2- Profil du candidat

Le terme candidat s'adresse aux candidates et candidats et nous gardons candidat par souci d'alléger le texte. Pour cette recherche, nous recherchons des candidats ayant une certaine expérience en programmation et algorithmique. Il serait intéressant que le candidat sache programmer dans au moins 2 de langages suivants : C/C++, python, perl, ruby. Également il doit être à l'aise dans la programmation en bash sous environnement GNU/Linux. Nous souhaitons également que le candidat ait ou soit disposé à étudier des algorithmes, packages et techniques d'intelligence artificielle et apprentissage profond. Le candidat devra par ailleurs posséder des connaissances en mathématiques (algèbre linéaire et optimisation principalement). Puisque la production de texte sera bilingue (français, espagnol), la connaissance de la langue espagnole est un atout.

3- Opportunités de mobilité à l'international du doctorant(e)

Oui, à l'Université Autonome Métropolitaine (UAM, Mexico), l'Université Nationale Autonome du Mexique (UNAM, Mexico); ou bien à l'Université de Montreal (Québec)

4- Références bibliographiques

(1) L Moreno, J-M Torres-Moreno, R Wedemann, H Boucheneb. FLE : A Fuzzy Logic Algorithm for Classification of Emotions in Literary Corpora. KDIR pp. 202-209, 2020

(2) Juan-Manuel Torres-Moreno. Résumé automatique de documents : Une approche statistique. Hermès Lavoisier, 2011, ISBN 978-2-7462-3212-9

(3) Abstraction ou hallucination ? Etat des lieux et évaluation du risque pour les modèles de génération de résumés automatiques de type séquence-à-séquence, Akani, Eunice, Favre, Benoit and Bechet, Frederic, Actes de la 29e Conférence sur le Traitement Automatique des Langues Naturelles. Volume 1, 6, 2022, Avignon, France

(4) Juan-Manuel Torres-Moreno, Marc El-Bèze, Frédéric Béchet, Nathalie Camelin, Fusion probabiliste appliquée à la détection et classification d'opinions, DEFT'09, Paris, France, 22 juin 2009, 15p.

(5) G Bel-Enguix, G Sierra, H Gómez-Adorno, J-M Torres-Moreno, J-G Ortiz-Barajas, J Vásquez. Overview of PAR-MEX at Iberlef 2022: Paraphrase Detection in Spanish Shared Task. Procesamiento de Lenguaje Natural (PLN),

(6) L Moreno, J-M Torres-Moreno. LiSSS: a new multi-annotated multi-emotional corpus of Literary Spanish Sentences, CyS, 24(3) :1139–1147, 20203

(7) L Moreno, J-M Torres-Moreno, R Wedemann. Generación de frases literarias: un experimento preliminar, PLN, 65 :29–36, 2020

(8) L Moreno, J-M Torres-Moreno, E SanJuan, R Wedemman. Automatic Generation of Literary Sentences, Linguamática, 12(1) :15-30, 2020 5

(9) L Moreno, J-M Torres-Moreno, C González. Estudio de hiperparámetros de modelos neuronales en la generación de frases literarias. Research in Computing Science (RCS), 150(5), 2021 20

(10) I Morgado, L Moreno, J-M Torres-Moreno, R Wedemann. MegaLite-PT: A Corpus of Literature in Portuguese for NLP. BRACIS 2022. Intelligent Systems pp 251–265

