

Séminaire DéLiCorTAL & Journée des doctorants (Axe 1)

Vendredi 14 février 2020 – Amphithéâtre
de la MSH – Saint Martin d'Hères

13:30 Conférence du Pr. Margarita ALONSO-RAMOS (Universidade da Coruña, Espagne) : *Le rôle du corpus dans l'apprentissage des collocations : corpus d'apprenants et corpus pour apprendre.*

14:30 Présentation de thèse 1 : Jirawan KIATPHOTHA (doctorante UGA, LIDILEM) : *Étude contrastive des expressions d'émotions formant des motifs textuels dans la littérature contemporaine française et thaï.*

15:00 Présentation de thèse 2 : Nicolas DAVID (doctorant UGA, LIDILEM) : *Approches méthodologiques dans la construction d'un corpus arboré pour le créole mauricien.*

Résumé de la conférence

Les collocations sont souvent considérées comme particulièrement difficiles pour les apprenants de L2, même pour les plus avancés. Cette affirmation apparaît généralement dans la littérature qui se concentre sur l'anglais L2, mais il apparaît nécessaire de vérifier si la même situation s'applique aux apprenants d'autres langues secondes, comme l'espagnol. Il est également nécessaire de confirmer si l'apprentissage basé sur les données (« data-driven learning ») qui est proposé pour améliorer la production de collocations par les apprenants d'anglais peut être transféré aux apprenants d'espagnol. Afin de vérifier cela, nous souhaitons répondre aux questions suivantes : 1) qu'est-ce qui caractérise l'utilisation des collocations par les apprenants d'espagnol en L2 d'après ce qu'on observe dans un corpus d'apprenants, et 2) le corpus peut-il aider les apprenants d'espagnol à améliorer leur production de collocations ? Dans cet exposé, je me concentrerai, premièrement, sur le travail que j'ai réalisé concernant l'annotation des collocations dans un corpus d'apprenants, et deuxièmement, sur le rôle du corpus de locuteurs natifs comme outil pour améliorer l'apprentissage des collocations. Je conclurai en présentant quelques réflexions sur le rôle pédagogique du corpus.

Résumé de la présentation de thèse 1

Le principal but de notre recherche est de contraster les expressions d'émotions en français avec celles en thaï. Ces expressions forment des motifs textuels qui présentent des régularités et des variations lexicales et syntaxiques ainsi que des fonctions discursives spécifiques (Legallois 2012, Longrée & Mellet 2013, Novakova et Siepmann 2020). Notre recherche s'appuie sur des approches fonctionnelles et contextualistes (Sinclair 2004, Hoey 2005, Biber 2009) qui nous permettent d'analyser les séquences relevées comme statistiquement spécifiques du corpus littéraire français à travers quatre niveaux d'analyse linguistique : lexical, sémantique, syntaxique et discursive. Les expressions sont annotées à l'aide d'une grille sémantique, créée dans le cadre du projet PhraseoRom (<https://phraseorom.univ-grenoble-alpes.fr>). Les expressions seront ensuite regroupées en motifs textuels par rapport à leur spécificité (calcul du Log-likelihood, fréquence) et par rapport à la dimension sémantique Qualia de la grille PhraseoRom avec ses trois valeurs : affect, sensation et perception (*j'ai honte, j'ai le vertige, je l'entends dire*). Notre étude sera fondée sur un corpus de littérature blanche contemporaine en français (PhraseoBase/Lexicoscope, 34 M. de mots) et un corpus comparable de textes littéraires contemporains en thaï, constitué pour les besoins de la thèse à partir du *Thai National Corpus* (8 M. de mots) et de textes littéraires contemporains en ligne (26 M. de mots). Il s'agit d'un sujet de thèse interdisciplinaire en phraséologie, linguistique de corpus et en linguistique contrastive.

Résumé de la présentation de thèse 2

L'intérêt du domaine du Traitement Automatique des Langues (TAL) pour les langues peu dotées ne cesse de croître, mais le manque de ressources linguistiques, d'outils informatiques et de corpus adéquats, auquel elles sont sujettes, tend à restreindre leur application et exploitation en TAL. Se situant ainsi dans une perspective de modélisation linguistico-informatique, l'objectif de ce projet de thèse est de procéder à la construction d'un corpus arboré pour une langue peu dotée : le créole mauricien. Dans un premier temps, cette présentation s'articulera autour des moyens employés afin de constituer un corpus écrit électronique. Puis, nous présenterons les principales phases de traitement du corpus en cours de constitution, et nous nous appesantirons sur l'importance et la nécessité du processus de normalisation. Par la suite, nous dresserons une esquisse du schéma d'annotation syntaxique étant envisageable, tout en exploitant les conventions orthographiques, lexicales et grammaticales de cette langue créole à base française.