



Nom – Prénom Fort Karën

Laboratoire de rattachement LORIA, équipe Sémagramme

Intitulé du diplôme HDR Informatique

Titre de l'HDR : Myriadisation et éthique pour le traitement automatique des langues

Abstract

Le traitement automatique des langues (TAL) a subi deux révolutions ces dix dernières années : le raccourcissement extrême de la distance entre les productions de la recherche et l'utilisateur final et l'avènement de l'apprentissage profond (*deep learning*). En conséquence, les besoins en données ont explosé en parallèle des questions éthiques. Cette habilitation à diriger des recherches présente les travaux que j'ai menés dans le domaine de la production d'annotations manuelles pour le TAL par myriadisation (*crowdsourcing*), en particulier par le jeu (*games with a purpose*), et dans celui de l'éthique pour le TAL. J'y redéfini la myriadisation et les sciences participatives en général et je présente en détail les jeux ayant un but, leurs atouts et leurs limites. Je m'attarde plus particulièrement sur ZombiLingo, qui a servi à collecter des annotations en syntaxe de dépendances pour le français et RigorMortis, un jeu d'annotation d'unités polylexicales. Je me concentre dans une dernière partie sur l'éthique pour le TAL, un sous-domaine qui n'a véritablement été reconnu qu'à partir de 2016 et dont j'ai été préceuseure. Je reviens sur son historique, son évolution récente et présente mes travaux, menés dans une optique plus déontologiste que conséquentialiste, permettant d'avoir une vision systémique du TAL et des problèmes éthiques qu'il pose.