



Documentation des langues et linguistique informatique : présentation de travaux en cours

ALEXIS MICHAUD

CNRS - LACITO

alexis.michaud@cnrs.fr

Résumé : Les méthodes informatiques présentent un fort potentiel pour aider à faire face à l'urgence que constitue la documentation des langues en danger. Les modèles créés par l'apprentissage machine peuvent aider à accomplir des tâches chronophages telles que la transcription et la traduction mot à mot (réalisation de gloses interlinéaires). Mais le traitement automatique des langues reste peu utilisé dans la documentation linguistique, notamment parce que la technologie est encore nouvelle (et en évolution rapide), et que la prise en main des outils demande des compétences spécialisées. L'exposé reviendra sur des travaux en cours dans trois directions, deux applicatives et la troisième plus fondamentale. La première consiste à créer une interface conviviale à l'usage des non-informaticiens. La seconde : mettre en place des équipes pluridisciplinaires qui puissent déployer, à l'échelle, les meilleurs outils, tirant parti du libre accès aux données favorisé par la transition en cours vers les pratiques de Science ouverte. La troisième direction explorée consiste à tirer parti des modèles statistiques au profit de la recherche fondamentale (en linguistique aussi bien qu'en informatique), en sondant les représentations neuronales.
