

L'extraction de connaissances spatiales à partir de documents textuels peut être une tâche difficile du fait de l'ambiguïté propre au langage naturel. L'indisponibilité de gros volumes de données étiquetées rend difficile la mise-en-œuvre d'un processus de découverte automatique. Dans ce contexte, nous abordons le problème de la désambiguïsation des entités spatiales, entre « localisation » et « organisation », par apprentissage actif. D'abord, nous introduisons une méthode de résolution des toponymes basée sur une analyse lexicale et contextuelle. Ensuite, nous proposons une amélioration en intégrant un modèle d'apprentissage actif. Celui-ci permet de sélectionner automatiquement les données non étiquetées les plus informatives pour l'annotation humaine. Les expérimentations sont réalisées sur un corpus de « SemEval-2007 » en anglais et soulignent l'amélioration du modèle d'apprentissage initial avec un étiquetage réduit.