

Constitution d'un lexique bilingue (Français, Arabe) à base d'un corpus parallèle

Abbaoui Houda, Tahri Soumia, Hoceini Youssef
Département Mathématique et Informatique
Faculté Sciences Exactes, Université Tahri Mohamed
Bechar, Algérie
yhoceini@gmail.com

Abstract—Ce travail s'inscrit dans le cadre du Traitement Automatique des Langues (TAL), plus précisément dans le domaine de la Traduction Automatique. La traduction automatique en tant que tel a besoin des ressources linguistique multilingue (lexique bilingue). L'utilité des corpus pour la lexicographie bilingue est mentionnée depuis certain temps et les projets de dictionnaire ont fait appel à des corpus parallèle tel que (Oxford-Hachette French Dictionary). L'extraction de ces lexiques bilingues faits à partir de corpus parallèle qui sont constitués de plusieurs volets correspondant chacun à une version d'un même texte dans deux langues différentes ou plus. Cette entité textuelle est un ensemble de versions électroniques de traductions et de leurs originaux respectifs, alignés au niveau des paragraphes, phrases ou même des mots. Dans notre application nous avons choisis un corpus parallèle formé d'un bi texte composé des arrêtés et décrets téléchargé à partir du site du ministère de l'enseignement supérieur et la recherche scientifique. La méthodologie adoptée dans la phase d'alignement (phrase à phrase et mot à mot) est bien celle d'un model linguistique qui repose sur l'ordre des mots dans les deux langues le français et l'arabe.

Keywords—Traitement automatique des langues, traduction automatique, corpus parallèle, alignement, lexique bilingue.