

# Système automatique d'analyse morpho-lexicale d'un Corpus Arabe

Belmaoun Hafsa, Hoceini Youssef  
Département Mathématique et Informatique  
Faculté des Sciences Exactes, Université Tahri Mohamed  
Béchar, Algérie  
yhoceini@gmail.com

*Résumé—A l'heure actuelle, le Traitement Automatique des Langues (TAL) et plus précisément la langue arabe, fait l'objet de nombreux travaux concernant d'une part, la modélisation linguistique propre à la langue, d'autre part, la conception et la réalisation des applications.*

*Notre projet s'intéresse à l'analyse morpho-lexicale d'un corpus arabe non voyellé, le processus de traitement adapté porte sur une approche d'analyse morpho-lexicale approfondie du contenu du corpus. La stratégie adoptée repose tout d'abord sur une phase de constitution du corpus dans un domaine tel que la culture, puis une autre de segmentation du corpus au niveau du texte, paragraphe, la phrase ainsi à l'intérieur du mot afin de décortiquer les différentes unités élémentaire (préfixe, suffixe, racine et schème) et enfin la conception et réalisation d'un module d'analyse à base de règles capable de faire : le prétraitement morpho lexical, l'analyse affixale, l'analyse morphologique et le post-traitement et enfin une catégorisation morpho-lexicale des différents types de mots arabes, à savoir, les noms, les verbes, et les particules (mots outils, mots spéciaux).*

*Mots-clés—Traitement automatique de l'arabe, corpus, Analyseur morpholexical, segmentation, règles, , analyse affixale, catégorisation morpho-lexicale.*