

EL.Bouyakhf¹, S.Sabri¹, A.Yousfi²

¹Faculté des Sciences, Université Mohammed V Agdal - Rabat

²Institut d'Etudes et Recherches pur l'Arabisation - Rabat

bouyakhf@fsr.ac.ma

sabri@fsr.ac.ma

yousfi71@yahoo.com

Le système d'analyse morphologique des noms dérivatifs arabes

Résumé:

Le but de cette présentation est d'exposer un système d'analyse morphologique des noms arabe qui utilise une nouvelle technique de traitement qui consiste à chercher le degré de similarité entre le mot entrée et la forme (wazne) sans faire appel à un corpus des mots. Le système présenté dans cette communication utilise une base de données de toutes les formes des noms dérivatifs arabes. L'algorithme d'analyse consiste à chercher dans la base la forme exacte du mot entrée, et à construire par la suite sa racine exacte en affichant certaines caractéristiques du mot entrée tels que les suffixes, les préfixes, le type et le genre

1. Introduction

L'analyse morphologique est une opération qui consiste à extraire des informations de types morphologiques et/ou syntaxiques du mot à analyser : la racine, la nature morphologique et syntaxique du mot, le temps de conjugaison, le pronom, les suffixes, les préfixes, etc. L'opération d'analyse morphologique automatique est réalisée à l'aide d'une application. Plusieurs travaux ont été réalisés dont le but d'élaborer des analyseurs morphologiques de la langue arabes, on cite à titre d'exemple :

- ✓ Hegazi et Elsharkawi qui ont développé un système pour extraire la racine d'un mot en utilisant le schème (wazn) morphologique de ce mot et sa nature [Hegazi 86] ;
- ✓ Saliba et Al-Dannan ont élaboré un système d'analyse et de génération morphologique pour le centre scientifique IBM au Kuwait. Cet analyseur permet de trouver toutes les analyses morphologiques possibles d'un mot [Saliba 89] ;
- ✓ El Sadany et Hashish ont développé un système d'analyse et de génération morphologique de la langue arabe. Ce système est capable de traiter les mots arabes voyellés, semi-voyellés ou non voyellés. Ce système est développé dans le centre scientifique du Caire [Sadany 89] ;
- ✓ Al-Fedaghi et Al-Anzi ont présenté un algorithme d'extraction de la racine et le schème d'un mot arabe. Cet algorithme est basé sur la détection de la position des lettres de la racine trilitère et quadrilitère pour vérifier son existence dans la liste des racines déjà connues [Fedaghi 89].
- ✓ Buckwalter a développé un système d'analyse morphologique de la langue Arabe en utilisant la technique de translittération Arabe/français, l'algorithme de l'analyseur consiste à déterminer toutes les segmentations possibles du mot puis chercher les résultats dans le corpus des radicaux, des suffixes ou des préfixes [Buckwalter].

2. Présentation de SAMNA

Le système d'analyse morphologique des noms dérivatifs arabes (SAMNA) est une application qui a été conçue et développée dans le but de fournir à partir d'un nom dérivatif conjugué, qui constitue son entrée, sa racine correspondante et ses différentes caractéristiques morphologiques. Le nom conjugué doit posséder une racine et obéir à une forme. La réalisation d'un tel travail nécessite :

- la construction d'une base de donnée contenant la plupart des formes des noms dérivatifs,
- la mise en place des techniques nécessaires permettant l'extraction des différentes caractéristiques morphologiques et grammaticales des noms conjugués à analyser.

3. Principe de fonctionnement de SAMNA

Le principe d'analyse consiste à chercher le degré de similarité entre la forme et le mot d'entrée. Cette tâche est accomplie à l'aide d'un module qui renvoie en résultat une distance D . L'idée originnaire de cette technique revient à Mr A. Yousfi [Yousfi-2004] lors de l'élaboration d'un Système d'Analyse Morphologiques des verbes, seulement il a supposé que la distance doit être maximale, alors pour que le système soit plus général (traite les verbes et les noms) et afin de trouver toutes les analyses possibles, il faut que cette distance soit supérieur à zéro ($D > 0$).

Le système utilise une base de données des formes des noms dérivatifs voyellées stockées dans un fichier texte qui est constitué de cinq champs (forme, racine du nom, genre, type du noms, racine verbal)

Le programme est constitué de deux modules : le premier permet de chercher la forme exacte du nom à analyser grâce à la distance alors que le deuxième est pour extraire des informations (racine, suffixes, préfixes...)



Figure 1 : Exemple d'analyse

4. Conclusion et perspectives

Des fonctionnalités peuvent être mises en place et rendre l'analyseur morphologique plus performant sans toutefois mettre en cause l'existant, afin de réduire le temps d'exécution.

Enfin, nous envisageons d'entreprendre cette étude pour développer un système d'analyse morphologique des verbes et des noms dérivatifs non voyellés et de la généraliser pour les tous les mots de la langue arabe.

5. Bibliographies.

1. [Fedagi 89] Al Fedaghi, S .S., and Al-Anzi, F.S. 1989 A new application to generate Arabic Root-Pattern Forms, Proceedings of the 11th National Computer Conference and Exhibition, March, Dahrn, Saudia Arabia, 391-400.
2. [Hegazi86] Hegazi, N., and ElSharkawi, A. A. 1986. *Natural Arabic Language Processing, Proceedings of the 9th National Computer Conference and Exhibition, Riyadh, Saudi Arabia, 1-17.*
3. [Saadani89] El-Sadany, T. A., and Hashish, M. A. 1989. *An Arabic Morphological System. IBM Systems Journal. Vol.28, No.4, 600-612.*
4. [Saliba89] Saliba, B., and Al-Dannan, A. 1989. *Automatic Morphological Analysis of Arabic : A Study of Content Word Analysis. Proceedings of the First Kuwait Computer Conference, Kuwait, March, 3-5.*
5. [Buckwalter] *Arabic Morphology Analysis* <http://www.qamus.org/>