

Perspectives

Dans le domaine de l'apprentissage artificiel les pistes de la recherche sont nombreuses. L'accouplement des techniques de cette dernière avec celles du traitement automatique de la langue pousse les frontières de ces pistes encore plus loin. L'obtention d'un doctorat est le début d'une carrière consacrée à la recherche et l'avancement scientifique. Les travaux menés dans cette thèse ne sont qu'un point de départ qui ouvre la voie envers d'autres perspectives qui méritent d'être explorées. Citons quelques exemples :

- **L'augmentation de la taille et du nombre des catégories dans le corpus :** les corpus utilisés dans les travaux menés dans cette thèse sont assez petits. La conception d'un plus grand corpus d'apprentissage permettra d'obtenir des résultats plus solides et d'attaquer d'autres problèmes provenant de la richesse de son contenu. Les travaux sur la classification automatique des documents écrits en caractères arabes d'un point de vue multilingue en sont un exemple : ce n'est qu'en examinant les documents des catégories « science et technologie » et « médecine » incorrectement classés que nous avons découvert cette piste très intéressante. Par conséquent, en augmentant le nombre de catégories et en y ajoutant de nouveaux documents des nouvelles perspectives s'ouvriront.
- **L'enrichissement de la base de données DIINAR.1 :** par exemple, il faut ajouter des informations permettant de traiter les mots empruntés ou translittérés. L'enrichissement de la liste des mots outils et l'intégration des informations permettant de traiter les entités nommés en sont d'autres. Tout cela permettra d'améliorer l'efficacité de l'analyse morphologique ainsi que la fidélité de la classification automatique.
- **L'expérimentation avec d'autres algorithmes d'apprentissage :** Nous nous sommes limités dans cette thèse à travailler avec quatre algorithmes d'apprentissage. Toutefois, le nombre d'algorithmes proposés dans la littérature est largement plus grand. Il est donc nécessaire de travailler avec d'autres algorithmes d'apprentissage.

- **La considération des techniques de classification automatique non-supervisée.** Tous les travaux menés dans cette thèse se basent sur la classification automatique supervisée. Les techniques issues de la classification automatique non-supervisées sont prometteuses. Comme c'est le cas avec l'approche dite supervisée, la littérature portant sur la classification automatique des documents écrits en caractères arabes non supervisée est pauvre, voire plus pauvre que l'approche dite supervisée. Pour cela, il faut considérer les techniques de la classification automatiques non-supervisée pour tester leur efficacité et leur validité en ce qui concerne la langue arabe.
- **La considération d'autres choix pour la nature des attributs :** Il est nécessaire de mener une étude plus élaborée sur le meilleur choix à faire pour la nature des attributs. Il faut considérer d'autres choix. Citons, à titre d'exemple, l'approche hybride où l'attribut est conçu à partir d'une analyse morphologique et puis des n-grammes ou vice versa surtout que l'analyse morphologique peut profiter de la richesse de la base de données DIINAR.¹