

ACHIT Abdelmadjid⁽¹⁾, AZZOUNE Hamid⁽²⁾

⁽¹⁾Laboratoire de Traitement Automatique de la Langue, Centre de Recherche Scientifique et Technique pour le Développement de la Langue Arabe CRSTDLA

aachit@yahoo.com

⁽²⁾Laboratoire de Recherche en Intelligence Artificielle LRIA, Département d'informatique, Faculté de Génie Electrique, Université Houari Boumediène USTHB

hazzoune@yahoo.fr

Comparaison des modélisations des niveaux d'analyse de l'arabe par les grammaires GPSG et HPSG

Résumé:

De nos jours, le traitement automatique de la langue arabe connaît des efforts de recherches considérables ayant pour objectif la conception et le développement de ressources linguistiques qui faciliteraient l'usage de l'arabe dans les différents domaines scientifiques et techniques. Et ce, en tirant profit des nouvelles technologies disponibles et des résultats scientifiques atteints dans le domaine du traitement automatique des langues naturelles TALN. Le développement et la mise en œuvre d'un linguisticien pour une langue donnée, dépend de l'application visée et les différents niveaux de traitements (morpho-lexical, syntaxique, sémantique et pragmatique) auxquels il se réfère, n'auront pas la même importance. Ainsi, certains niveaux seront favorisés par rapport à d'autres. Cependant, le gros du travail, demeure la formalisation et la modélisation des connaissances linguistiques de cette langue, de manière à faciliter l'automatisation des traitements portant sur ses textes. Jusqu'à aujourd'hui, les travaux de recherches en intelligence artificielle IA, en général, et en TALN, en particulier, menés dans le cadre de la formalisation des structures syntaxiques des langues naturelles essayaient pour la plupart de proposer des outils et formalismes ad hoc et des fois dépassant ce stade là. Ce qu'il y'a lieu de noter, c'est que tous travaux se sont heurtés à la formidable complexité de la langue naturelle qui semble bien résister à toutes ces tentatives. Et peu d'entre eux, ceux qui arrivent à assurer une large couverture linguistique des différents phénomènes linguistiques rencontrés dans une langue naturelle. La multitude de formalismes et d'outils linguistiques existants, souleva le problème du choix du formalisme qui convient le mieux à la langue traitée. Dans cette quête, nous devons être guidés par quelques critères dont nous citons: la facilité d'implémentation, la déclarativité, l'universalité des principes, la large couverture linguistique, etc.

Dans cet article, nous présentons une étude portant sur la comparaison des modélisations des niveaux d'analyse de l'arabe par les grammaires syntagmatiques généralisées GPSG et les grammaires syntagmatiques guidées par les têtes HPSG, en se basant sur les concepts de la théorie linguistique Néo-khalilienne TNK. L'objectif visé, au travers de cette étude, est de sélectionner sur la base de critères bien définis, lequel de ces deux formalismes qui sera adopté pour servir de plateforme à la conception d'outils TAL pour la langue arabe. Le choix des grammaires d'unification n'est pas fortuit. Il découle du fait des caractéristiques attrayantes qu'ont cette famille de grammaires comparées aux autres catégories de grammaires. L'étude comparative nécessita la définition et l'établissement de critères de comparaison que nous avons organisé en deux grandes catégories: catégorie dite des critères généraux et catégorie dite des critères dépendants de la langue étudiée.

En dernier, cette étude aboutit au fait que HPSG, comme formalisme représentant les différents niveaux d'analyse de l'arabe, est à plus d'une raison recommandé et préconisé que GPSG. Ainsi, il a été établi que HPSG propose un lexique beaucoup plus riche et mieux organisé que GPSG. De plus, concernant la représentation des constructions syntaxiques arabes, HPSG offre une représentation plus élégante, plus générale et plus économique en terme de coût : type et nombre de règles utilisées.

Mots-clés: traitement automatique de la langue arabe, analyse syntaxique, modélisation linguistique, structure de traits, grammaires d'unification, GPSG, HPSG