

## Construction d'une grammaire d'arbres adjoints pour la langue arabe

Fériel Ben Fraj

- (1) Laboratoire RIADI, École Nationale des Sciences de l'Informatique, 2010 Manouba, Tunisie.  
Ferial.BenFraj@riadi.rnu.tn

**Résumé.** La langue arabe présente des spécificités qui la rendent plus ambiguë que d'autres langues naturelles. Sa morphologie, sa syntaxe ainsi que sa sémantique sont en corrélation et se complètent l'une l'autre. Dans le but de construire une grammaire qui soit adaptée à ces spécificités, nous avons conçu et développé une application d'aide à la création des règles syntaxiques licites suivant le formalisme d'arbres adjoints. Cette application est modulaire et enrichie par des astuces de contrôle de la création et aussi d'une interface conviviale pour assister l'utilisateur final dans la gestion des créations prévues.

**Abstract.** The Arabic language consists of a set of specificities. Thus, it is more ambiguous than other natural languages. Its morphology, syntax and semantic are correlated to each other. We have constructed an application for the construction of a tree adjoining grammar which respects the characteristics of the Arabic language. This tool allows constructing the grammatical rules as elementary trees enriched by different feature structures. It helps the user by its interface and control system to manage correct and uniform rules.

**Mots-clés :** Outil semi-automatique, grammaire d'arbres adjoints, langue arabe, traits d'unification  
**Keywords:** semi-automatic tool, tree adjoining grammar, Arabic language, feature structures

### 1 Introduction

La langue arabe présente des caractéristiques qui la rendent plus ambiguë que d'autres langues naturelles, spécialement les langues indo-européennes. Pour cette langue, il y a un entrelacement fort entre les différents niveaux du traitement linguistique ; à savoir morphologique, syntaxique et aussi sémantique. Ainsi, la syntaxe ne peut être bien gérée qu'avec l'intervention de données à la fois morphologiques et aussi sémantiques. Cette corrélation se manifeste, par exemple, dans l'accord en genre et en nombre entre les noms et les adjectifs ou encore l'accord entre les noms et les verbes. Ces accords dépendent, intimement, de l'animation et/ou de l'humanité du nom en question. Citons alors l'exemple des phrases suivantes : الكلاب جائعة (*les chiens sont affamés*) : le mot الكلاب (*les chiens*) étant un nom masculin pluriel inhumain ne s'accorde pas en nombre et en genre avec l'adjectif جائعة (*est affamée*) qui est au féminin singulier. Ceci n'est pas le cas avec la phrase الأطفال جوع (*les enfants sont affamés*), où l'accord entre le nom et l'adjectif existe, vu que الأطفال (*les enfants*) est un nom animé humain, et ainsi l'adjectif جوع (*sont affamés*) est au masculin pluriel. D'autant plus, le schème (الوزن) de l'item lexical qui est une information morphologique aide à spécifier l'utilisation syntaxique de cet outil voire même son sens, tel par exemple le schème فاعل (*Fa'ilun*) qui indique le réalisateur de l'action.

Par conséquent, la construction d'une grammaire formelle pour la langue arabe doit, obligatoirement, faire intervenir différents types d'informations utiles afin de mieux gérer la bonne composition des agencements syntaxiques. Dans ce papier, nous dressons le problème de cette construction. Il existe différentes grammaires formelles ; à savoir : les grammaires génératives (Chomsky, 1965) et celles

transformationnelles (Lecomte, 2004) de Chomsky, les grammaires d'unification (Kahane, 2002) et aussi les grammaires à la fois génératives et d'unification, comme par exemple la grammaire de clauses définies (DCG) (Pereira et Warren, 1980). Parmi toute cette panoplie, nous avons choisi d'utiliser le formalisme grammatical d'arbres adjoints (TAG : *Tree Adjoining Grammar*) (Joshi et al., 1975). C'est un formalisme d'unification qui manipule les règles grammaticales sous forme de structures arborescentes dites « arbres élémentaires » pouvant être reliées entre elles par des opérations de jointure bien définies dans (Joshi et al., 1975). La justification de ce choix a été discutée dans l'article (Ben Fraj et al., 2008). Néanmoins, nous n'étions pas les premiers à choisir ce formalisme. Nous devons mentionner qu'un travail antérieur a été élaboré par (Habash et Rambow, 2004) pour générer une grammaire TAG pour la même langue par extraction des règles (ou arbres élémentaires) à partir d'un corpus arboré (Maamouri et al., 2004). Pour notre part, nous visons une construction, plutôt théorique qui touche une granularité informationnelle assez riche et fine à la fois, et ce dans le but d'obtenir des arbres élémentaires exacts, uniformes et à large couverture. La grammaire, ainsi, construite constitue une adaptation de TAG aux spécificités de l'arabe (Ben Fraj, 2010).

Le présent papier est structuré comme suit. La seconde section présente un état de l'art des grammaires formelles déjà construites pour la langue arabe. Nous nous y focalisons sur les grammaires d'unification étant donné que c'est le même courant que nous suivons. La troisième section est réservée à la description de la grammaire que nous visons construire ; à savoir ses caractéristiques, ses arbres élémentaires ainsi que les structures de traits qui y sont incarnées. Une quatrième section présentera l'application développée pour assister l'utilisateur final à bien gérer les agencements syntaxiques. Une évaluation succincte de l'outil présente la première version de la grammaire construite. Nous concluons cet article par la présentation d'un ensemble de perspectives concernant l'amélioration de l'outil développé et l'utilisation de notre grammaire.

## 2 Grammaires pour la langue arabe : Etat de l'art

La construction d'une grammaire présente une tâche primordiale pour différentes applications linguistiques. Ainsi, pour la langue arabe, différentes grammaires suivant différents formalismes grammaticaux ont été bâties. Le but principal est de servir la tâche d'analyse syntaxique. Dans cette section, nous présentons, brièvement, les différentes grammaires qui ont été conçues pour représenter la syntaxe de l'arabe. Tous ces travaux constituent des tentatives pour couvrir la plupart des règles syntaxiques de la langue arabe ainsi que les phénomènes linguistiques assez riches qui leurs sont attachés.

La TAG a été déjà utilisée pour la représentation de la syntaxe arabe. (Habash et Rambow, 2004) ont construit une grammaire d'arbres adjoints par extraction d'arbres élémentaires à partir d'un corpus arboré. Pour ce faire, ils ont utilisé la partie 1 de la version 2.0 de PATB (*Penn Arabic TreeBank*) (Maamouri et al., 2004). Cette partie contient 160,000 mots annotés qui composent des textes du style journalistique. L'extraction passe par une réinterprétation du corpus en des structures de dépendances. Les chercheurs ont pu extraire des structures avec variation des positions de leurs composants vu que la langue arabe est à ordre semi-fixe. Ils ont, alors, obtenu des phrases à composition VSO (Verbe Sujet Objet) et d'autres à SVO, alors qu'ils ont échoué à obtenir la composition VOS. Cet échec est dû à l'absence de ce genre de structures dans le corpus utilisé pour l'extraction (Habash et Rambow, 2004). Le problème de couverture est posé vu que cette construction est liée aux styles de textes contenus dans le corpus.

La grammaire LFG (*Lexical Functional Grammar*) (Kaplan et Bresnan, 1982), a aussi été utilisée pour représenter la syntaxe de la langue arabe. Un travail a été élaboré par Attia (Attia, 2008). La grammaire qui a été créée met en relief différentes spécificités de la langue arabe, spécialement les phrases nominales ; c'est-à-dire les phrases sans verbes, les phrases à sujet caché, les phrases nominales à propos (خبر) phrase verbale, etc. Il a, aussi, spécifié les équations d'accord entre les différents composants des structures syntaxiques. Pour toute phrase représentée par cette grammaire, les deux structures (la *c-structure* et la *f-structure*) de la LFG ont été construites.

La HPSG (*Head driven Phrase Structure Grammar*) (Pollard et Sag, 1987) a fait aussi l'objet d'une recherche pour la description de la syntaxe arabe. Cette recherche a été élaborée au sein du laboratoire LARIS (Haddar et al., 2010). L'idée de base consiste à construire une grammaire qui possède suffisamment de contraintes afin de l'utiliser pour la tâche d'analyse syntaxique. Les chercheurs ont essayé de couvrir plusieurs phénomènes linguistiques de la langue manipulée. Ils ont, alors, construit les structures des phrases simples et aussi celles modifiées (par des mots outils) ainsi que les structures

représentant des syntagmes nominaux simples et aussi à compositions récursives (comportant des subordinées relatives).

### 3 ArabTAG : une adaptation de TAG à la langue arabe

La grammaire que nous avons conçue est baptisée ArabTAG (*Arabic Tree Adjoining Grammar*). Elle hérite tous les fondements de base de TAG qui ont été accommodés aux spécificités de la langue arabe. Notre grammaire décrit hiérarchiquement différents composants syntaxiques de différents niveaux : phrases, syntagmes et mots (Ben Fraj et al., 2008) ainsi que les différentes informations qui leurs sont relatives (morphologiques, syntaxiques et aussi sémantiques). ArabTAG admet aussi un ensemble de caractéristiques que nous décrivons succinctement, dans ce qui suit.

#### 3.1 ArabTAG est générique

En ArabTAG, seules les structures syntaxiques génériques sont représentées ; c'est-à-dire les structures qui sont formées par seulement les éléments nécessaires à leur bonne formation. Ces éléments sont des sous-catégories syntaxiques (syntagme, verbe, adjectif, etc) affectés par des rôles qu'ils jouent au sein de la structure mère où ils co-occurrent (Abeillé, 1993). Ainsi, pour une phrase verbale, par exemple, ce sont : le verbe, le sujet et le(s) complément(s) d'objet qui sont mis en valeur. Dans le cas d'une phrase nominale, المبتدأ (*le thème*) et الخبر (*le propos*) sont les éléments à représenter. Il y a aussi description des structures génériques correspondant aux syntagmes nominaux et ceux prépositionnels. Les compléments de ces composants syntaxiques, comme par exemple ظرف الزمان (*le C.C. de temps*) et ظرف المكان (*le C.C. de lieu*) ou encore الحال (*le mode*), peuvent être liés aux arbres élémentaires d'ArabTAG en utilisant les opérations appropriées d'association de structures (Joshi, 1975). Le but de cette généralité consiste à éviter l'explosion du nombre des arbres élémentaires possibles qui risquent d'être indénombrables, sachant que la langue arabe est à ordre semi-fixe (Blachère et Gaudefroy-demombynes, 1975). L'explosion est inévitable si nous prenons en considération toutes les compositions possibles des éléments syntaxiques, par variation de l'ordre des éléments co-occurents qu'ils soient obligatoires ou facultatifs. Par conséquent, ArabTAG renforce la notion de partage de structures, qui constitue un des avantages de TAG.

#### 3.2 Richesse des structures de traits

Dans ArabTAG, nous avons défini différents types d'informations sous forme de traits : morphologiques, syntaxiques, sémantiques et compositionnels. Tous ces traits sont nécessaires pour la bonne gestion des compositions syntaxiques. Ils sont organisés au sein de structures dites « traits d'unification ». Chaque nœud, au sein d'un arbre élémentaire, possède sa propre structure de traits lui permettant de gérer ses associations ultérieures possibles avec d'autres arbres (Joshi, 1975). A chaque catégorie de nœuds est associé un ensemble précis de traits. A chacun de ces traits, nous avons défini un ensemble de valeurs qui décrivent les relations du composant syntaxique cible avec son contexte. Le tableau 1 englobe un ensemble exemple et non exhaustif des traits d'unification.

Trait	Description	Nœuds concernés
<i>Gen</i>	Spécifie l'accord en genre.	N, Adj, Dem ou SN
<i>Nombr</i>	Précise l'accord en nombre.	N, Adj, Dem ou SN
<i>Cas</i>	Définit le type du cas : Nominatif, Accusatif, Génétif ou Gémiation.	N, Adj ou SN
<i>TestCas</i>	Vérifie l'accord en cas.	N, Adj, SN
<i>Def</i>	Spécifie l'obligation (+) ou non (-) de la définition de l'item cible	N, SN
<i>TypeDef</i>	Deux types de définition en arabe : définition par l'article "الـ" (A) ou par annexion (B).	N, SN
<i>TestDef</i>	Vérifie la nécessité (+) ou non (-) d'un accord du nom cible en définition avec son contexte.	N
<i>Pronom</i>	Pronom = Genre+Nombre. Spécifie l'accord entre le verbe et son sujet (resp. suppléant du sujet <sup>1</sup> ) en genre et/ou en nombre,	V

1. Le suppléant du sujet (نائب فاعل) remplace le sujet dans les phrases de la voix passive

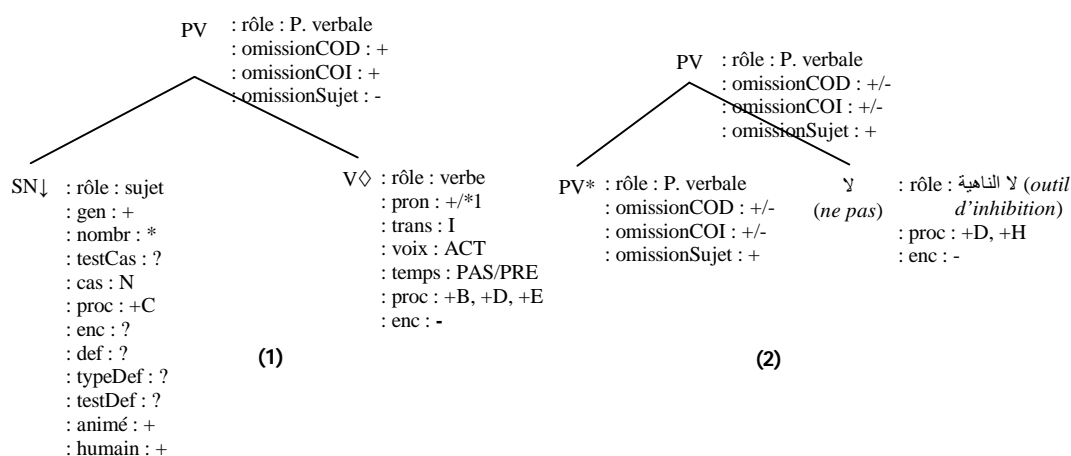
<i>Trans</i>	Vérifie la nécessité de l'existence des compléments d'objet ( <b>TI</b> , <b>TD</b> , <b>I</b> ).	V
<i>Voix</i>	Indique la voix de conjugaison du verbe.	PV, V
<i>Temps</i>	Indique le temps de conjugaison du verbe.	V
<i>Proc</i>	Caractérise les types de proclitiques permis pour l'item cible.	Tous les nœuds
<i>Enc</i>	Spécifie si le composant syntaxique nécessite (+) ou non (-) d'être agglutiné à un enclitique.	Tous les nœuds
<i>OmissionTheme</i>	Définit une obligation (+) ou non (-) de l'omission du thème d'une phrase nominale.	PN
<i>OmissionSujet</i>	Définit une obligation (+) ou non (-) de l'omission du sujet.	PV
<i>OmissionCOD</i>	Définit une obligation (+) ou non (-) de l'omission du COD.	PV
<i>OmissionCOI</i>	Définit une obligation (+) ou non (-) de l'omission du COI.	PV
<i>OmissionComp</i>	Définit une obligation (+) ou non (-) de l'omission du complément du nom dans un SN d'annexion (dans le cas où le nom à compléter est agglutiné à un enclitique).	SN-Annexion
<i>Propre</i>	Spécifie si le nom est propre ( <b>P</b> ) ou commun ( <b>C</b> ).	N, SN
<i>Rôle</i>	Le rôle du composant courant au sein de l'arbre élémentaire.	Tous les nœuds
<i>Animé</i>	Indique si l'élément linguistique est animé (+) ou non (-).	N et tous les SN
<i>Humain</i>	Indique si l'élément correspond à un être humain (+) ou non (-).	N et tous les SN

Tableau 1 : Présentation des traits d'unification dans ArabTAG

Dans ce tableau les symboles miss en gras entre parenthèses correspondent à des exemples de valeurs associées aux différents traits. En plus de ces traits d'unification, nous avons défini un second ensemble : celui des traits d'instanciation. Ce sont des informations morphosyntaxiques et sémantiques nécessaires lors de la lexicalisation de la grammaire.

### 3.3 ArabTAG est partiellement lexicalisée

Une grammaire d'arbres adjoints lexicalisée est une grammaire dans laquelle tout arbre élémentaire est muni d'au moins un symbole terminal qui est un item lexical. ArabTAG n'est pas complètement lexicalisée, mais elle l'est partiellement. L'objectif de ce choix consiste, encore une fois, à factoriser au maximum les structures élémentaires. De ce fait, nous mentionnons qu'ArabTAG contient un ensemble d'arbres élémentaires lexicalisés réservés pour représenter les contextes possibles des mots outils jouant le rôle de « modificateurs » de phrases et/ou de syntagmes. En outre, notre grammaire possède d'autres structures non lexicalisées, appelées aussi « modèles d'arbres ». Ces modèles comportent des nœuds dits « ancrés ». Ces nœuds sont susceptibles de recevoir des entités lexicales. La figure 1 présente ces deux types d'arbres.



<sup>2</sup> Les valeurs du trait proc présentent les classes des proclitiques pouvant s'agglutiner aux mots (+D : حروف العطف (proclitiques de coordination)).

Dans cette figure, alors que les traits correspondants aux racines sont compositionnels, les autres permettent plutôt la gestion des correspondances d'informations à l'intérieur de la structure du composant syntaxique. Dans l'arbre (1), nous mettons l'accent sur les valeurs à associer aux traits. Si l'on commence à partir de la racine, nous remarquons que les attributs *omissionCOD* et *omissionCOI* sont à « + » ; c'est-à-dire que ces deux omissions sont obligatoires vu que le verbe dans la structure est intransitif « I ». Pour la même raison, le trait *enc* du verbe est à « - » ; c'est-à-dire que ce verbe ne doit pas être agglutiné à un enclitique. Pourtant, l'omission du sujet dans cette structure n'est pas permise car un sujet existe réellement dans la structure. Le symbole « ? » veut dire que la valeur est indéfinie à ce stade et elle peut l'être dans les dérivations suivantes du composant. Pour le genre du côté du SN, sa valeur (+) veut dire qu'il y a obligatoirement un accord en genre avec le verbe. Concernant la valeur du nombre (\*), l'accord est conditionné par le fait que le verbe soit conjugué avec un pronom singulier ou non. Ceci nous mène à exploiter la valeur associée au pronom du côté verbe (« +/\*1 »). Cette quantité veut dire que l'accord en genre est obligatoire et que la conjugaison du verbe doit être seulement réalisée avec des pronoms singuliers: هو (*Il*) ou bien هي (*Elle*). Cette règle d'accord est basée sur les valeurs des deux traits sémantiques : animé et humain. Lorsque ces deux traits sont actifs (« + »), le verbe au début de la phrase doit être toujours au singulier et seul son genre change. Cette structure peut représenter les deux phrases suivantes: نام الطفل (*L'enfant s'est endormi*) et لعبت البنت (*La fille a joué*). Le second arbre (2) de la figure constitue un arbre élémentaire lexicalisé et auxiliaire (Joshi, 1975). Elle met l'accent sur les traits compositionnels qui dépendent fortement de la composition interne de la phrase verbale (PV).

## 4 Présentation de l'outil de construction d'ArabTAG

### 4.1 Description

Dans le but de construire ArabTAG tout en respectant ses caractéristiques, déjà décrites, nous avons choisi de concevoir et développer une application informatique qui consiste à aider un grammairien de la langue arabe lors de la construction des arbres élémentaires d'ArabTAG. Cet outil est à la fois modulaire et enrichi d'un système de contrôle du processus de création. De ce fait, il permet de garantir l'exactitude et l'uniformité des règles (ou arbres élémentaires) construites. Ladite application présente une interface conviviale qui élabore la structure hiérarchique de l'arbre.

La construction d'un arbre élémentaire passe obligatoirement par la construction de ses différents nœuds qui détiennent toutes les informations nécessaires pour sa bonne formation. La construction des nœuds se fait selon un mode en profondeur. Ainsi, un nouveau nœud ne peut être construit que si son prédécesseur (son nœud frère) est complètement créé avec tous ses niveaux hiérarchiques internes prévus. Le nœud racine est le premier à être construit. La boîte de dialogue à droite de la figure 2, permet d'attribuer à cette racine les traits compositionnels qui lui sont relatifs. Ces traits autorisent ou refusent à la racine d'admettre certains composants internes comme le montre l'exemple (1) de la figure 1. Cette même boîte de dialogue affecte à chaque nœud l'ensemble qui lui correspond de traits d'unification. Cet ensemble varie selon que le type du nœud (nominal, verbal ou particule). L'affectation des traits et de leurs valeurs aux nœuds est contrôlée par l'outil. Ainsi, selon le type du nœud, le bon ensemble de traits est activé alors que les autres traits sont désactivés.

D'autant plus, un arbre peut accepter un (ou plusieurs) nœud(s) ancre(s). Tout nœud ancre admet, en plus des traits d'unification, le second ensemble de traits que nous venons de nommer d'instanciation. L'ensemble de ces traits sont illustrés dans la boîte de dialogue à gauche de la figure 2. Cette boîte est appelée lors de la création d'un nœud ancre au sein d'une structure élémentaire générique ou encore lors de la création d'un arbre lexicalisé. Elle permet d'associer des traits divers de type morphosyntaxiques et sémantiques au nœud en question.

Chaque information ajoutée à la structure élémentaire est représentée d'une façon littéraire sur l'interface principale de l'application (au milieu de la figure 2), et ce en attente de son codage à la fin de la construction avec le métalangage XML. Nous avons choisi XML dans le but de bien structurer les données contenues dans la grammaire et de standardiser leur représentation et ce afin de les exploiter par la suite aisément et indépendamment des applications.

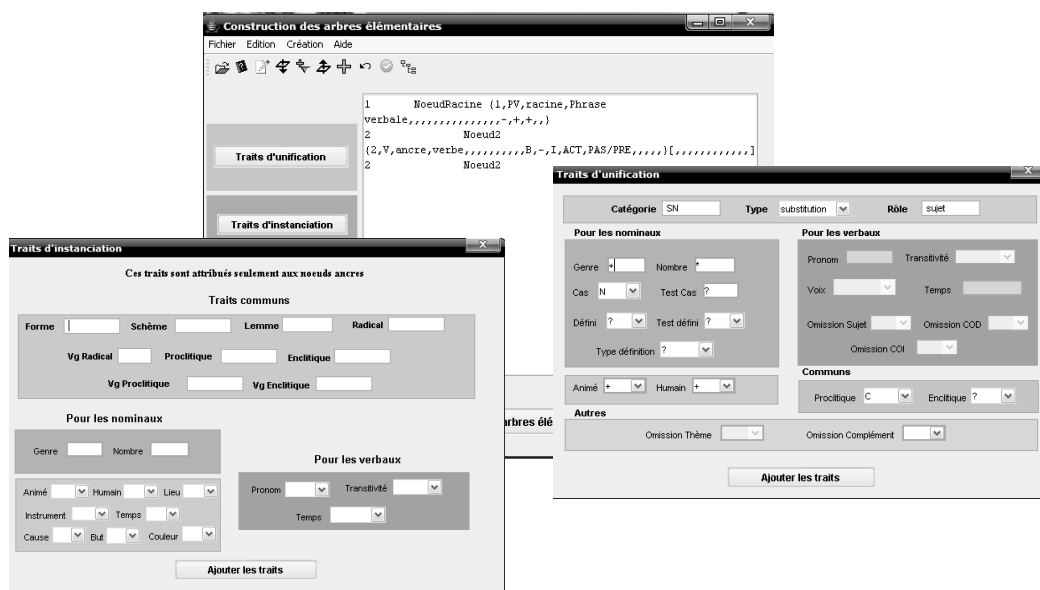


Figure 2 : Différentes interfaces de l'application de construction d'ArabTAG

#### 4.2 Evaluation de l'outil : la version 1.0 d'ArabTAG en chiffres

L'application, étant construite, nous a permis de construire une première version d'ArabTAG. Elle est composée de différentes structures syntaxiques réparties entre phrases et syntagmes. Nous avons tenu compte de la diversité des structures syntagmatiques, spécialement nominales. Nous avons, alors, obtenu 380 arbres élémentaires différents, répartis entre quatre grandes familles comme le montre la figure 3.

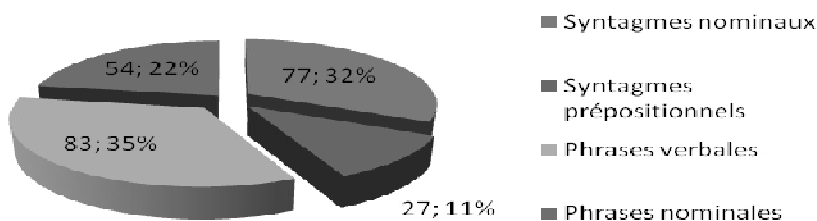


Figure 3 : Répartition des structures élémentaires d'ArabTAG en familles de structures

Lors de la création de cette version d'ArabTAG, les bases théoriques que nous avons suivies sont les livres de Dahdah (Dahdah, 1992) et de Kouloughli (Kouloughli, 1992). La grammaire, ainsi construite couvre des structures elliptiques, d'autres anaphoriques et aussi des structures renfermant des subordonnées. Elle prend en considération la variation des positions des éléments au sein des composants syntaxiques et le phénomène d'agglutination.

#### 5 Conclusion et perspectives

Afin de construire une grammaire d'arbres adjoints pour la langue arabe, nous avons choisi de développer un outil pour assister et aider les linguistes à élaborer des arbres élémentaires corrects, suffisamment riches et uniformes. L'application est à interface conviviale et elle possède des astuces pour contrôler la procédure de construction les règles sur le plan hiérarchisation ou sur le plan structures de traits. Suite à l'utilisation de cet outil, nous avons pu obtenir la version 1.0 de notre grammaire contenant 380 arbres élémentaires. Toutefois, notre outil peut être amélioré par ajout d'un système de prédiction pour la création qui se base sur la version, déjà construite, d'ArabTAG pour mieux assister l'utilisateur à mener les constructions ultérieures. D'autant plus, la version construite n'est que première et peut facilement être alimentée par de nouveaux arbres élémentaires, comme elle peut être transformée en une version complètement lexicalisée. La grammaire, ainsi obtenue, peut être utilisée dans différentes applications

linguistiques, spécialement l'analyse syntaxique de la langue arabe, et ce étant donnée la richesse de ses structures. Notre outil d'aide à la construction peut aussi être amélioré par d'autres modules comme ceux de conversion de la TAG à d'autres formalismes grammaticaux.

## Références

ABEILLE A. (1993). *Les nouvelles syntaxes*, Armand Colin.

ATTIA MOHAMMED A. (2008). Handling Arabic Morphological and Syntactic Ambiguity within the LFG Framework with a View to Machine Translation, Thèse de doctorat, Faculty of humanities, University of Manchester.

BEN FRAJ F. (2010). Un analyseur syntaxique pour les textes en langue arabe à base d'un apprentissage à partir de patrons d'arbres syntaxiques, Thèse de doctorat, Ecole Nationale des Sciences de l'Informatique, Université de La Manouba.

BEN FRAJ F., BEN OTHMANE ZRIBI C., BEN AHMED M. (2008). Ensemble Classification for Parsing Arabic Texts: A Theoretical Approach, *In Proceedings of the AIPR 2008 multi-conference*, Orlando, Florida, USA, 51-57.

BLACHÈRE R., and GAUDEFRY-DEMOMBYNES M. (1975). *Grammaire de la langue arabe*, Maisonneuve et Larose.

CHOMSKY N. (1965) *Aspects of the theory of syntax*, Cambridge, Massachusetts: MIT Press.

DAHDAH A. (1992). *معجم قواعد اللغة العربية في جداول ولوحات*, Librairie de Nachirun Lebanon, 5ème edition, 1992.

HABASH N., RAMBOW O. (2004). Extracting Tree Adjoining Grammar from the Penn Arabic Treebank, *JEP-TALN 2004, Session Traitement Automatique de l'Arabe*, Fès.

HADDAR K., BOUKEDI S., ZALILA I. (2010). Construction of an HPSG grammar for the Arabic language and its specification in TDL, *International Journal on Information and Communication Technologies*, vol. 3, n° 3.

JOSHI A. K., LEVY L. S., TAKAHASHI M. (1975). Tree adjunct grammars, *Journal of the Computer and System Sciences*, vol. 10, n° 1, 136-163.

KAHANE S. (2002). Grammaire d'Unification Sens-Texte – Vers un modèle mathématique articulé de la langue, Hdr de l'Université Paris 7, 82.

KAPLAN R. M., BRESNAN J. (1982). Lexical functional grammar: A formal system for grammatical representation, *In J. BRESNAN, Ed., The Mental Representation of Grammatical Relations*. Cambridge, Mass.: MIT Press, 173-281.

KOULOUGHLI J. (1992). *La grammaire Arabe pour tous*, Edition Presses Pocket.

LECOMTE A., *Méthodes pour le Traitement Automatique des Langues*, cours Ingénierie de la Communication Personne-Système, Université Pierre Mendès-France, France, 2004.

MAAMOURI M., BIES A., BUCKWALTER T., MEKKI W. (2004). The Penn Arabic Treebank: Building a Large-Scale Annotated Arabic Corpus, *In NEMLAR International Conference on Arabic Language Resources and Tools*, 102-109.

PEREIRA F. C. N. and WARREN D. H. D. (1980). *Definite clause grammars for language analysis : a survey of the formalism and a comparison with augmented transition networks*. Artificial Intelligence 13, pp. 231 – 278.

POLLARD C., and SAG I. (1994). Head-driven Phrase Structure Grammar, *CLSI series*, University of Chicago.