

Modélisation de l'ordre des mots en arabe standard

Dina El-Kassas, Sylvain Kahane

Lattice – Université Paris VII
dina@linguist.jussieu.fr, sk@ccr.jussieu.fr

Résumé – Abstract

Cet article propose une description de l'ordre des mots pour le verbe et ses compléments en arabe standard en postulant la présence systématique du verbe en position initiale.

This paper proposes a description of the word order for the verb and its complements in Standard Arabic by postulating the systematic presence of the verb in initial position.

Keywords – Mots Clés

Syntaxe, arabe standard, ordre des mots, modèle topologique, grammaire de dépendance.
Syntax, standard Arabic, word order, topological model, dependency grammar.

Introduction

La première partie de notre article est une analyse linguistique des différents ordres des mots dans la phrase déclarative en arabe. Nous considérerons tout particulièrement l'ordre SVO. Cet ordre pose problème, car la forme du verbe y est différente de celle de l'ordre standard VSO. Nous postulerons la présence d'un auxiliaire vide (une forme de *kaana* 'être' au présent) en tête de la construction SVO.

Les deuxième et troisième parties esquisseront une implémentation de notre analyse dans le cadre de la syntaxe de dépendance et du modèle topologique. Cette approche considère la structure syntaxique d'une phrase comme un arbre de dépendance non ordonné ; l'ordre des mots et le regroupement des mots en constituants sont capturés par une autre structure, que nous appelons la structure topologique (Gerdes & Kahane 2001).

1 L'ordre des mots en arabe standard

L'ordre standard en arabe standard est l'ordre VSO (Al-Chartouni 1986, Wright 1859) :

- (1) *ʔakala* *al#ʔawlaad+u* *al#tuffaah+a* أكلَ الأولادُ التفاحَ
(V)PAST.3.MASC DEF#(N,masc)PL+NOM DEF#(N,masc)SG+ACC
a/ont mangé les enfants les pommes
'les enfants ont mangé les pommes'

Le verbe possède une flexion temporelle et s'accorde en personne et genre avec le sujet. Le sujet est au nominatif et l'objet direct à l'accusatif. Leur ordre respectif est libre (ordre VSO ou VOS). L'objet direct peut être soit topicalisé (2a), soit disloqué (2b) (ordre OVS). L'objet disloqué est au nominatif et est repris par un pronom clitique tandis que la topicalisation consiste juste à déplacer l'objet à l'accusatif dans une position initiale :

- (2) a. *al#tuffaah+a* *ʔakala* *al#ʔawlaad+u*
 b. (*al#tuffaah+u*) *ʔakala#hu* *al#ʔawlaad+u*
 DEF#(N,masc)SG+NOM (V)PAST.3.MASC#(PRO,3)MASC.SG.ACCDEF#(N,masc)PL+NOM
 ‘(les pommes,) les enfants les ont mangés’

Le sujet peut également précéder le verbe (ordre SVO). Dans ce cas, la flexion du verbe est remplacée par un élément que la grammaire traditionnelle analyse comme un pronom clitique sujet¹ :

- (3) (*al#ʔawlaad+u*) *ʔakal#uu* *al#tuffaah+a* الأولادُ أكلوا التفاحَ
 DEF#(N,masc)PL+NOM (V)PAST#(PRO,3)MASC.PL.NOM DEF#(N,masc)SG+ACC
 les enfants ont mangé les pommes
 ‘(les enfants,) ils ont mangé les pommes’

Néanmoins l'exemple (3) est problématique, car, s'il s'apparente à une dislocation du sujet, il est possible avec une prosodie sans détachement de l'élément antéposé. De plus, cet ordre est très productif, notamment dans le dialecte égyptien, que nous avons plus particulièrement considéré. Nous allons proposer pour (3) une analyse particulière qui nécessite un détour par les constructions avec *kaana* ‘être’, également très productives.

- (4) *kaana* *al#ʔawlaad+u* *ja ʔakul#uuna* *al#tuffaah+a*
 (AUX)PAST.3.MASC DEF#(N,masc)PL+NOM (V)PRESENT#(PRO,3)MASC.PL.NOM DEF#(N,masc)SG+ACC
 étais(en)t les enfants mangés les pommes
 ‘les enfants étaient en train de manger les pommes’

La phrase précédente est une phrase du type « *kaana* SVO » où l'on retrouve pour S, V et O les mêmes formes qu'en (3). Le groupe nominal *alʔawlaadu* (S) est le sujet de *kaana* avec lequel il s'accorde, *ja ʔakul#uuna* (V) ayant un sujet pronominal coréférent avec S. Nous considérerons qu'il s'agit de la juxtaposition d'une propositions P1 = *kaana* N et d'une proposition P2 avec un pronom coréférent avec N. Ce pronom est préférentiellement sujet, mais pas obligatoirement.

Cette construction avec juxtaposition est fortement grammaticalisée² et l'ensemble se comporte comme une seule clause avec une coprédication entre *kaana* et le verbe V de

¹ Une autre analyse considère l'arabe comme une langue pro-drop (= à sujet pronominal vide) et la deuxième forme du verbe comme une forme verbale avec accord dit fort en genre, personne et nombre. Ce point n'étant pas pertinent pour notre étude, nous adopterons le point de vue traditionnel même si celui-ci est par ailleurs problématique.

² Notons une autre construction avec juxtaposition également très grammaticalisée : l'arabe a seulement un passif morphologique sans agent, mais la réalisation de l'agent peut être assurée par la juxtaposition d'une proposition de même verbe à l'actif : التفاحة أكلت، أكلها زيد lit. : la pomme a été mangée, Zayd l'a mangée, 'la pomme a été mangée par Zayd'.

P2 comme le montre la modalité temps-aspect de P1P2. En effet, celle-ci résulte de la combinaison des grammèmes de temps sur *kaana* et V de la façon suivante. Considérons une séquence $kaana_{\alpha} S V_{\beta}$ où α et β sont l'un des trois grammèmes de temps (passé, présent, futur). Le grammème α exprime la position temporelle par rapport à « maintenant » (le temps absolu : ‘passé’, ‘présent’, ‘futur’), tandis que β exprime le temps relatif par rapport à α et a donc une valeur aspectuelle (‘accompli’, ‘en train d’avoir lieu’, ‘à venir’). Par exemple, $kaana_{\text{futur}} S V_{\text{passé}}$ se traduira en français par un futur antérieur (سيكون الأولاد قد أكلوا التفاح *les enfants auront mangé*), tandis que $kaana_{\text{passé}} S V_{\text{futur}}$ se traduira par un conditionnel passé (كان الأولاد سيأكلون *les enfants auraient dû manger*).

Ce tableau est compliqué par le fait que la forme présent de *kaana* n'est jamais utilisée (dans la construction précédente comme dans la construction où *kaana* est copule). Nous postulons donc que *kaana* a une forme vide au présent et que l'ordre SVO (comme en (3)) est, quand S n'est pas disloqué, une occurrence de $kaana_{\text{présent}} SVO$.

Nous pensons ainsi que les constructions $V_{\text{passé}} S$ et $S V_{\text{passé}}$ ne sont pas tout à fait synonymes : $V_{\text{passé}} S$ exprime un passé pur (أكل الأولاد *les enfants mangèrent*), tandis que $S V_{\text{passé}}$ (= $kaana_{\text{présent}} S V_{\text{passé}}$) exprime un présent accompli (الأولاد أكلوا *les enfants ont mangé*), à distinguer encore de $kaana_{\text{passé}} S V_{\text{présent}}$ qui exprime une action en train d'avoir lieu dans le passé (كان الأولاد يأكلون *les enfants mangeaient/étaient en train de manger*).

Nous allons maintenant proposer une syntaxe de dépendance pour l'arabe (section 2) et donner une description de l'ordre des mots dans le modèle topologique (section 3).

2 Syntaxe de dépendance de l'arabe

Nous représentons la *structure syntaxique* d'une phrase par un arbre de dépendance (non ordonné) : les nœuds de l'arbre sont étiquetés par les mots de la phrase (ou plus exactement leur décomposition morphologique) et les branches par les relations syntaxiques entre mots. Le verbe principal est la racine de l'arbre. La figure 1 représente la structure syntaxique de la phrase (4) :

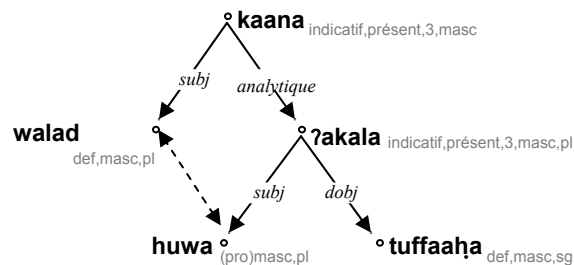


Figure 1

La figure 2 représente la structure syntaxique de la phrase (2b), où l'élément disloqué dépend du verbe par une relation proleptique (*mubtada?*):

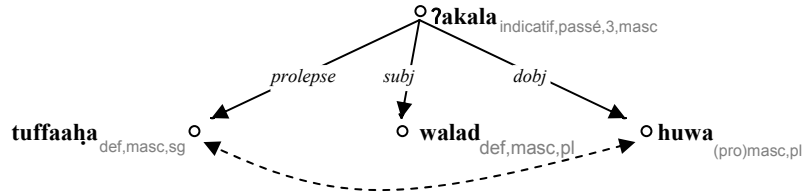


Figure 2

La structure SVO, illustrée en (3), a deux analyses, suivant la prosodie. La première analyse consiste à considérer qu'il s'agit d'un ordre communicativement neutre avec pour tête syntaxique l'auxiliaire *kaana* au présent, « *kaana*_{présent}SVO » (figure 3a). La deuxième analyse consiste à considérer qu'il s'agit d'une dislocation du sujet avec reprise pronominale (figure 3b).

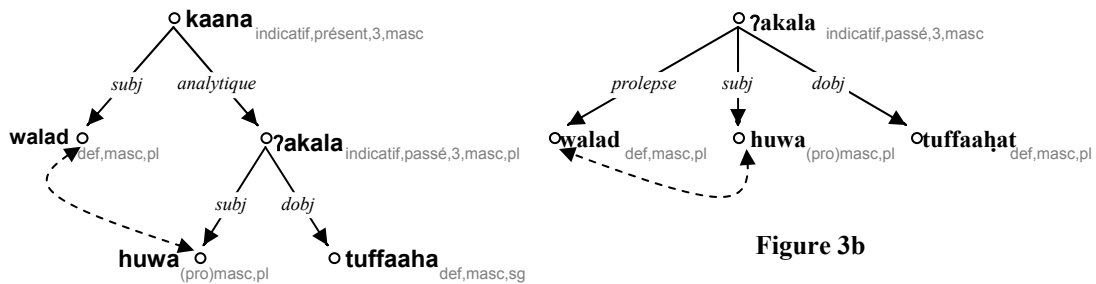


Figure 3a

Figure 3b

3 Un modèle topologique pour l'arabe

A côté de la structure syntaxique proprement dite, nous considérons une structure dite *topologique* qui décrit l'ordre des mots et leur groupement en constituants. Les constituants sont vus comme des boîtes comportant une liste de champs dans lesquels vont venir se placer d'autres boîtes. La grammaire topologique qui décrit la correspondance entre la structure syntaxique et la structure topologique d'une phrase comprend plusieurs types de règles (cf. Gerdes & Kahane 2001) :

- Une *règle de correspondance* décrit dans quel champ F peut être placé un mot de catégorie syntaxique C_2 dépendant d'un mot de catégorie syntaxique C_1 par une relation syntaxique r ; la règle précise également quelles frontières de boîte le mot dépendant peut franchir pour atteindre le champ F, ce champ pouvant *a priori* appartenir à n'importe quelle boîte contenant le mot gouverneur.
- Une *règle de description de boîte* donne la liste des champs ($F_1 \dots F_n$) d'une boîte B (et s'apparente à une règle de réécriture d'une grammaire hors-contexte).
- Une *règle de description de champ* indique combien de boîtes peuvent occuper un champ F.
- Une *règle de création de boîte* indique quelle boîte B peut être créée lorsqu'un mot de catégorie C est placé dans un champ F et dans quel champ F' de B le mot est placé.

Conformément à nos résultats de la première partie, nous considérons que la phrase déclarative arabe a pour tête syntaxique le verbe qui crée une première boîte, appelée *domaine principal* domP et composée de trois champs (chI chR chF). Le champ initial chI peut accueillir un nombre quelconque de constituants : il s'agit des éléments disloqués ou topicalisés. Le champ final chF accueille les éléments disloqués à droite. Le champ de rection chR accueille le verbe, qui ouvre le *domaine verbal* proprement dit, domV, comportant trois champs (chV chM chP). Le champ du milieu chM accueille les compléments non prédicatifs du verbe (sujet compris). Le champ prédicatif chP accueille optionnellement un complément prédicatif, qui ouvrira à nouveau un domaine verbal s'il s'agit d'un verbe, par exemple quand le verbe principal est *kaana*, le coprédicat ira dans le chP. Le champ verbal chV accueille le verbe, qui ouvre un *noyau verbal* noyV comportant six champs (chQ chN chS chL chC1 chC2). Les trois premiers champs accueillent un certain nombre de marqueurs adverbiaux : le marqueur d'énonciation *qad* (chQ), le marqueur de négation (chN) et le marqueur de futur *sawfa* (chS) ; le champ lexical chL accueille le verbe et les champs clitiques chC1 et chC2 accueillent successivement les clitiques sujet et objet. Seul le champ chL contient obligatoirement un élément, les autres peuvent être remplis ou rester vides.

Ces premières règles permettent de décrire les structures topologiques. Nous donnons dans la figure 4 la structure topologique de la phrase (4) (dont les structures syntaxiques sont données figure 1 et 2) et dans la figure 5 celle de la phrase (2b).

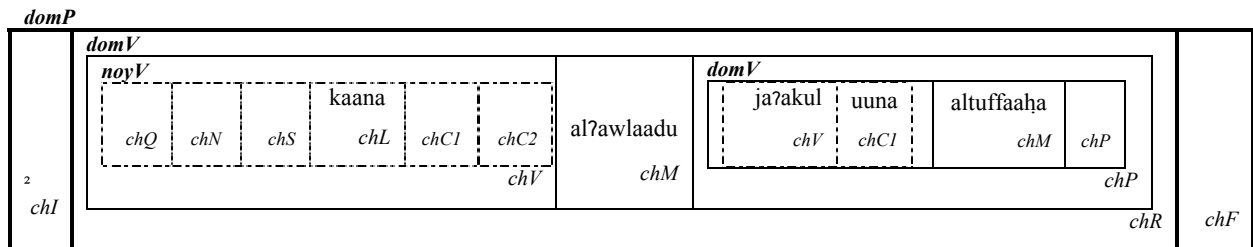


Figure 4: Structure topologique pour une construction *kaanaSVO*

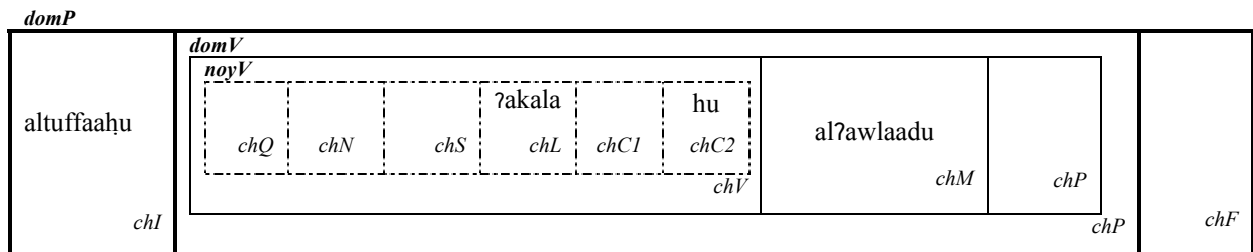


Figure 5: Structure topologique pour une dislocation gauche de l'objet direct

Nous allons maintenant donner les règles de correspondance de notre fragment de grammaire de l'arabe et montrer comment procéder au placement des éléments de la phrase et passer des structures syntaxiques des phrases à leurs structures topologiques. L'avantage de notre approche est que ces règles sont très simples bien qu'elles couvrent un fragment conséquent de l'arabe.

Les compléments non prédicatifs et non pronominaux du verbe (subj, dobj, ...) vont dans le chM du domV ouvert par leur gouverneur (et ne peuvent donc franchir la frontière d'un domV). Les pronoms sujet et objet vont dans les chC1 et chC2 du noyV ouvert par leur

gouverneur (et ne peuvent donc franchir la frontière d'un noyV). Le prolepse va dans le chI ou le chF ; le choix dépendra de la structure communicative. Un complément prédicatif va dans le chP de son gouverneur. Les marqueurs qui se placent devant le verbe vont dans les champs prévus à cet effet. Les autres dépendants du verbe ne sont pas traités ici.

Conclusion

Notre grammaire est implémentée dans le système DepLin développé par Kim Gerdes (www.talana.jussieu.fr/~kim/deplin) pour lequel des grammaires de l'allemand (Gerdes 2002), du coréen (Yoo 2003) et du français sont en développement. Nous avons montré que le verbe *kaana* joue un rôle central dans la grammaire arabe, notamment en permettant une double expression des grammèmes de temps à la base d'un système astucieux de temps-aspect. De plus en postulant une réalisation vide de *kaana* au présent, nous avons pu fournir une explication raisonnable à l'alternance des formes verbales dans les constructions VSO et SVO et proposer un modèle topologique simple de la phrase arabe.

Remerciements

Nous tenons à remercier Nor El-Hoda Arbaoui, Laurence Danlos, Adil El-Ghali, Kim Gerdes et tout particulièrement Igor Mel'cuk, pour leurs précieux commentaires.

Références

Al-Chartouni R. (1986), *mabaadi ʔu alʔarabijati*, Beyrouth, Dar El-Machreq.

Gerdes K., Kahane S. (2001), *Word order in German: A formal dependency grammar using a topological hierarchy*, ACL 2001, Toulouse.

Gerdes K. (2002), *Topologie et grammaires formelles de l'allemand*, thèse de doctorat, Paris, Université Paris VII.

Mel'cuk I. (1988), *Dependency Syntax: Theory and Practice*, The SUNY Press, N.Y.

Wright W. (1859, 1862, reprinted 1988), *A Grammar of the Arabic Language*, Cambridge, Cambridge University Press, 3^{ème} édition révisée.

Yoo, H. (2003), *Ordre des mots et prosodie : Essai de description et de formalisation pour le grec moderne*, thèse de doctorat, Paris, Université Paris VII.