Towards a Logical Approach to Nominal Sentences Analysis in Standard Arabic

Houda Anoun

Towards a Logical Approach to Nominal Sentences Analysis in Standard Arabic

Houda Anoun
Bordeaux 1-LaBRI-Signes (Inria Futurs)
anoun@labri.fr

Abstract. Standard Arabic (**SA**) is an extremely rich natural language that has unfortunately received very little interest within computational linguistics literature. We propose in this paper to explore this fertile ground and show the first steps towards the formalization of Arabic syntax and semantics by means of MultiModal Categorial Grammars. We will particularly focus on the analysis of some phenomena related to nominal sentences construction in **SA** using relevant packages of lexically anchored structural rules.

1 Introduction

Standard Arabic (**SA**) is an extensively used Semitic language: it is considered as the official language of more than twenty countries.
Although **SA**'s grammar has been studied since the 8^{th} century by Sibawayh (Sibawayh 1983) among others, very little thorough research work has been done on the formalization of its syntax and semantics. Moreover, all previous studies related to this field either focus on the syntactic level (e.g., using minimalist program (Abdel 2005; Kremers 2003)), or use linguistic models whose syntax/semantics interface is rather ad-hoc (e.g., using CFG (Haddad 2005)).
To the best of our knowledge, no attempt has yet been made to capture Arabic syntactic and semantic phenomena within a logical setting. We propose in the following paper to explore this promising direction and describe a fragment of **SA** by means of MultiModal Categorial Grammars (**MMCG**) (Moortgat 1997). Our initiative is fruitful since it allows us to take advantage of the transparent interface between syntax and semantics guaranteed by Curry-Howard correspondence.
In this paper, we will focus on the analysis of some nominal sentences phenomena in **SA** (e.g., word order, annexation phenomenon). We will particularly show how to capture such phenomena in an elegant manner using constrained structural reasoning.
Our survey aims at consolidating the interrelation between logic and natural languages which undoubtedly has a dual benefit. On the one hand, the use of a rigorous formalism such as **MMCG** will help us study **SA** linguistics in a neat and precise fashion. On

Proceedings of the Eleventh ESSLLI Student Session
Janneke Huitink & Sophia Katrenko (editors)

the other hand, the success of this formalization will confirm the linguistic relevance of **MMCG** model which proves to be readily adaptable to the specificities of a rich language such as **SA**.

2 Preliminaries

2.1 What is MMCG?

MMCG (Moortgat 1997; Oehrle 2001) is a logical system well suited to natural language analysis. This model proved relevant as it allows a neat analysis of complex linguistic phenomena occurring in various natural languages (e.g., Dutch verb clusters (Moortgat 1999), multiple Wh-questions in Serbo-Croatian (Vermaat 2005)).
MMCG are composed of two distinct parts: a constant one containing the core logic and a variable one which allows a controlled management of resources.
The first part is represented by a deductive system that describes invariants of grammatical form and meaning composition. Furthermore, it is equipped with a compositional construction process for semantics (represented using a simply typed λ-calculus) owing to Curry-Howard correspondence. The underlying logical rules of **MMCG** are universal as they do not depend upon the words (i.e., grammar terminals) of the chosen natural language. They rather express the way in which such words can combine by using their syntactic types. **MMCG** handles families of binary type constructors ($/_i$, $\backslash_i$, $\bullet_i$) provided with the structural counterpart $(,)^i$ and a set of unary connectives ($\Diamond_j$, $\Box_j$) associated with the structure-forming operator $<>_j$. The categorial slashes represent directional forms of the linear implication, they are used to express grammatical incompleteness. For instance, definite adjectives in **SA** (e.g., *al-mufīd-u*, i.e., interesting) take the type $\mathbf{np}\backslash_0\mathbf{np}$ to express their need to combine with a noun phrase to their left to form a modified noun phrase. However, English adjectives (e.g., interesting) take rather the type $\mathbf{n}/_0\mathbf{n}$ as they will merge with a common noun to their right to yield a modified noun. Moreover, unary operators can be used to encode various features such as morphosyntactic ones (e.g., case, gender, number etc) (Heylen 1999). Indeed, $\Box_j\mathbf{A}$ (resp. $\Diamond_j\mathbf{A}$) can be seen as a subtype of A with feature **j**. For example, we can assign the improved type $\Box_{sg}\Box_{ma}$ $\mathbf{np}\backslash_0\mathbf{np}$ to the adjective '*al-mufīd-u*' to explicitly specify that it requires a singular masculine noun phrase. Thus, we are able to avoid the analysis of some ungrammatical phrases such as **(*al-qiṣat-u al-mufīd-u*, i.e., the interesting-[mas] story-[fem]) which stems from the combination between the previous adjective and a feminine noun phrase.
The second part of **MMCG** encapsulates cross-linguistic variation by means of structural rules which allow controlled reconfigurations of contexts. Structural reasoning is constrained thanks to the use of modes of composition that play a crucial role within this framework (Oehrle 2001). Hence, instead of considering a globally available commutativity rule, we can assume that this property is verified by a particular family of connectives marked with mode **c**. We can control even more this local commutativity by restricting its application to configurations whose left sub-trees are decorated with the structural con-

nective $<>_j$. This latter structural rule is formalized by means of the following rewriting rule which can be applied to any appropriate sub-context during a given derivation:

$$P\diamond(c,j) : (<\Delta_1>_j, \Delta_2)^c \rightarrow (\Delta_2, <\Delta_1> j)^c$$

For the sake of legibility, we will rather present structural rules using their axiomatic form as shown below:

$$P\diamond(c,j) : \diamond_j A \bullet_c B \rightarrow B \bullet_c \diamond_j A$$

Structural reasoning is a powerful tool; it can be used to capture the flexibility of word-order in a neat fashion (Vermaat 2005). Indeed, it allows to relate the different structural positions that a word can occupy within a phrase, thus limiting its lexical ambiguity. Consequently, the application of relevant structural rules makes it possible to derive all the plausible configurations of a given clause from a single type assignment describing the canonical behavior of each one of its components. This asset will be used subsequently to account for word-order variation within **SA** nominal sentences.

In this paper, we use the natural deduction presentation of **MMCG**. For the purpose of completeness, the logical rules of this system are presented in Figure 1.1. We recall that the deduction rules operate on sequents like (Γ ⊢ x : A), where Γ is a structured binary context, **A** is a syntactic category and **x** is a simply typed λ-term that encapsulates the derivational semantics. The interested reader can find an in-depth survey of this deductive system in (Moortgat 1997).

$$\frac{}{x : A \vdash x : A}\ Ax \qquad \frac{\Delta \vdash p : A \bullet_i B \quad \Gamma[(a:A, b:B)^i] \vdash y : C}{\Gamma[\Delta] \vdash y[a := \Pi_1(p), b := \Pi_2(p)] : C}\ \bullet_i E \qquad \frac{\Gamma \vdash f : A/_i B \quad \Delta \vdash b : B}{(\Gamma, \Delta)^i \vdash (fb) : A}\ /_i E$$

$$\frac{\Gamma \vdash b : B \quad \Delta \vdash f : B\backslash_i A}{(\Gamma, \Delta)^i \vdash (fb) : A}\ \backslash_i E \qquad \frac{\Gamma \vdash a : A \quad \Delta \vdash b : B}{(\Gamma, \Delta)^i \vdash (a, b) : A \bullet_i B}\ \bullet_i I \qquad \frac{(\Gamma, x : B)^i \vdash f : A}{\Gamma \vdash \lambda x.f : A/_i B}\ /_i I$$

$$\frac{(x : B, \Gamma)^i \vdash f : A}{\Gamma \vdash \lambda x.f : B\backslash_i A}\ \backslash_i I \qquad \frac{\Delta \vdash x : \diamond_j A \quad \Gamma[<a : A>_j] \vdash c : C}{\Gamma[\Delta] \vdash c[a := x] : C}\ \diamond_i E \qquad \frac{\Gamma \vdash a : \Box_j A}{<\Gamma>_j \vdash a : A}\ \Box_j E$$

$$\frac{<\Gamma>_j \vdash a : A}{\Gamma \vdash a : \Box_j A}\ \Box_i I \qquad \frac{\Gamma \vdash a : A}{<\Gamma>_j \vdash a : \diamond_j A}\ \diamond_j I \qquad \frac{(\Delta_1 \overset{R}{\rightarrow} \Delta_2)\Gamma[\Delta_2] \vdash x : C}{\Gamma[\Delta_1] \vdash x : C}\ S_R$$

Figure 1.1: Natural deduction rules for Multimodal Logic

2.2 Some Words about SA

We present below two important characteristics of **SA** which are likely to help with the comprehension of our study. More details about **SA** grammar can be found in (Blachère 1994; Ryding 2005; Arrajihi 1975).

- **SA** is a highly inflectional language: cases are generally marked by means of suffixes (e.g., -*u*: nominative, -*a*: accusative and -*i*: genitive). Moreover, definiteness indicators are incorporated within nouns. In fact, the prefix *al*-[1] is used to form definite nouns while a suffix -*n* marks indefinite ones. For instance, *walad-u-n* (i.e., a boy) is an indefinite nominative noun, whereas *al-walad-i* (i.e., the boy) is a definite genitive one.

- **SA** is a language with mixed word-order: in fact, word-order in **SA** can be very flexible in some constructions (e.g., all of SVO, VOS, OVS and VSO orders are generally plausible) but so strict in others (e.g., adjectives always follow their modified nouns). It will be interesting to use controlled structural reasoning to deal with this diversity.

SA writing is built upon a specific alphabet and its direction is from right to left, but for the sake of readability, we will rather use the transliteration given by *arabtex*[2] package.

3 Syntax & Semantics of Nominal Sentences in SA

3.1 Basic Nominal Sentences Analysis

In contrast with languages such as English or French, we can build *nominal* sentences in **SA** that contain no verb (there is no copulative verb in Arabic such as '*to be*' or '*to remain*'). This construction is frequently used in other Semitic languages, notably in Hebrew (e.g., *ha-sepr gadol*, i.e., the book is big).
Nominal sentences give descriptions or definitions which are independent of time. They are composed of two components namely a *topic* realized by a noun phrase with nominative case and a *comment* which can be either an indefinite noun modifier, an indefinite noun or a prepositional sentence. Examples of grammatical and ungrammatical basic nominal sentences in **SA** are shown below:

(1) *al-manṭiq-u/*manṭiq-u-n* *mufīd-u-n*
(the) logic-[nom]/logic-[nom] interesting-[ind]
'Logic is interesting'

(2) *mufīd-u-n* *al-manṭiq-u*
interesting-[ind] logic-[nom]
'Logic is interesting'

(3) *al-walad-u* *fī 'l-bayti*
the boy-[nom] in the house
'The boy is in the house'

[1]The prefix *al*- becomes '*l*- when the noun occurs after a word ending with a vowel
[2]Available at www.informatik.uni-stuttgart.de/ifi/bs/research/arab_e.html

(4) *fī 'l-bayti 'l-walad-u*
in the house the boy-[nom]
'The boy is in the house'

(5) *fī 'l-bayti walad-u-n, *walad-u-n fī 'l-bayti*
in the house a boy
'A boy is in the house'

A definite topic (cf. ex. 1-4) can be placed either before or after its comment. Hence, the canonical order puts the emphasis on the definite topic (ex. 1&3), while the inverse order makes it possible to underline the comment (ex. 2&4). However, an indefinite topic can only be used with prepositional comments and should be placed at the end of the sentence (cf. ex. 1&5). In fact, the predicate which occurs after an indefinite noun is considered as an attributive adjective rather than a comment; the resulting construction is than a nominal phrase (of type **np**) instead of a nominal sentence (of type **s**[3]).
To account for the previous constraints that manage word-order between the topic and its comment in **SA**, we use controlled structural reasoning. Firstly, we assign to each constituent a single refined type which describes its canonical syntactic behavior and encapsulates its relevant morphosyntactic features[4] by means of $\Box$ operator:

Definite Topic	Indefinite Topic	Prep-Comment	Other Comments
al-manṭiq-u, al-walad-u	*manṭiqu-n, walad-u-n*	*fī 'lbayti*	*mufīdu-n*
$\Box_{def}\Box_{nom}$ np	$\Box_{ind}\Box_{nom}$ np	$s/_{c'}\Box_{nom}np$	$\Box_{def}\Box_{nom}$ $np\backslash_c$ s

Composition modes *c* and *c'* used in this lexicon are governed by the postulates below:

$$P(c)\text{: A} \bullet_c \text{B} \longrightarrow \text{B} \bullet_c \text{A} \quad P\Diamond(c', def)\text{: } (\Diamond_{def}\text{A}) \bullet_{c'} \text{B} \longrightarrow \text{B} \bullet_{c'} (\Diamond_{def}\ \text{A})$$

We notice that both modes *c* and *c'* are used to add local commutativity to our system. The commutativity introduced by mode *c* makes it possible to combine two expressions whose respective order is unconditionally free. In that case, both type constructors $/_c$ and $\backslash_c$ represent the same connective namely the non associative linear implication. On the other hand, mode *c'* introduces commutativity in a constrained fashion thanks to the use of the control operator $\Diamond_{def}$. Indeed, its associated structural rule, $P\Diamond(c', def)$, cannot be applied unless the first combined expression is definite.
If we consider the grammar provided with the lexicon and postulates above, then we are able to derive the correct examples and predict the ill-formedness of the ungrammatical ones in a straightforward manner. In fact, the underlined idea is as follows. When the comment is a noun modifier (or an indefinite noun), it combines with its definite topic using a commutative mode *c*, thus allowing the envisaged free word-order. On the other hand, a prepositional comment always searches for its topic to the right and combines with it using mode *c'*. As the definiteness of the topic required by the prepositional comment is

[3]In all this paper, the atomic type **s** represents well-formed nominal sentences
[4]Each np is decorated by its definiteness feature (*def*, *ind*) followed by its case feature

underspecified, a potential topic cannot enter the derivation until its definiteness feature is checked (i.e., by means of $\Box$E rule). If the topic is definite (i.e., at this stage, it should be decorated by the structural operator $<>_{def}$), then it can move to the beginning of the sentence thanks to $P\Diamond(c', def)$ postulate; otherwise, no displacement proves to be possible. These steps are illustrated in the derivation of sentence 3 below:

$$\dfrac{\dfrac{\dfrac{}{f\bar{\imath}'lbayti \vdash f : s/'_{c}\Box_{nom}np}\,Ax \qquad \dfrac{\dfrac{}{alwaladu \vdash w : \Box_{def}\Box_{nom}np}\,Ax}{< alwaladu >_{def} \vdash w : \Box_{nom}np}\,\Box_{def}E}{(f\bar{\imath}'lbayti, < alwaladu >_{def})^{c'} \vdash (fw) : s}\,/_{c'}E}{(< alwaladu >_{def}, f\bar{\imath}'lbayti)^{c'} \vdash (fw) : s}\,P\Diamond(c', def)$$

The derivational semantics of a sentence is computed in tandem with its syntactic derivation thanks to Curry-Howard correspondence. For instance, the deduction associated to sentence 3 yields the term (f w), where f (resp. w) represents the semantics of *al-walad-u* (resp. *fī 'lbayti*). The final semantics of this sentence, namely the logical formula **in(**ι **man,** ι **house)**[5], results from substituting each formal variable representing a linguistic entity by its lexical semantics.

3.2 Towards Complex Nominal Sentences

Annexation Phenomenon

All topics so far have been simple noun phrases. We will see in this section how we can enhance our nominal sentences by using compound topics.
In **SA**, we can form compound noun phrases by means of *annexation* phenomenon (Blachère 1994; Kremers 2003). These compound nouns have the following form 'cn=n_1 n_2 ... n_k' (k $\geq$ 2), where each n_j (1≤j<k) is a noun in *construction state* (i.e., which has neither the definite nor the indefinite indicator), whereas n_k is a noun phrase (either definite or indefinite). The resulting compound noun 'cn' inherits the definiteness feature from n_k, whereas its case is the same as n_1 (all the other nouns n_j (j≥2) take the genitive case). Here are some examples of noun phrases built using annexation:

(6) *ʾibn-u/*al-ʾibn-u* *'l-mudarris-i*
son-[nom]/the son-[nom] the teacher-[gen]
'the son of the teacher'

(7) *kitāb-u/*kitāb-u-n* *mudarris-i-n*
book-[nom]/a book-[nom] a teacher-[gen]
'the book of a teacher'

[5] Recall that ι, the description operator, is of type (e→t)→e and (ι P) returns the only individual that verifies the property P

(8) *kitāb-u* *ʾibn-i* *'l-mudarris-i*
book-[nom] son-[gen] the teacher-[gen]
'the book of the teacher's son'

To capture annexation phenomenon within **MMCG**, we assign a suitable syntactic type to each one of the three classes of **SA** nouns, namely al-nouns (definite simple nouns), cs-nouns (nouns in construction state) and nn-nouns (indefinite simple nouns).

al-nouns	nn-nouns	cs-nouns
$\Box_{al}\Box_{case}$np	$\Box_{nn}\ \Box_{case}$np	$\Box_{cs}(\ \Box_{case}\text{np}\ /_0\ \Box_{gen}\ \text{np})$

Hence, al-nouns and nn-nouns are both complete and self-contained as they can be used in several contexts (as subjects or topics etc). Nevertheless, cs-nouns are incomplete; they are only used to build compound noun phrases. They are assigned a functional type since they require to combine with a noun phrase to their right by means of a rigid composition mode 0 (i.e., non-associative and non-commutative mode) to yield a complete expression. We consider the grammar that supports the following package of postulates, $\mathcal{R}$= K(*def*) ∪ K(*ind*) ∪ Inc(*def*, *al*) ∪ Inc(*ind*, *nn*), where:

$$\text{K(j)}:\ \Diamond_j(\text{A} \bullet_0 \text{B}) \longrightarrow \Diamond_{cs}\ \text{A} \bullet_0 \Diamond_j\ \text{B} \quad \text{Inc(i, j)}:\ \Diamond_i\text{A} \longrightarrow \Diamond_j\text{B}$$

The structural rule K(*def*) (resp. K(*ind*)) is a kind of strong distributivity postulate (Heylen 1999). Intuitively, this postulate stipulate that a complex constituent is definite (resp. indefinite) if its head is in construction state (e.g., cs-noun) and its complement is definite (resp. indefinite). However, the rule Inc(*def*, *al*) (resp. Inc(*ind*, *nn*)) is nothing else but an inclusion principle. It states that all al-nouns (resp. nn-nouns) are inevitably definite (resp. indefinite).

Owing to the package $\mathcal{R}$, only well-formed compound nouns '$n_1 ... n_k$' can be derived, they are assigned either type '$\Box_{def}\Box_{c_1}$np' if n_k is an al-noun or type '$\Box_{ind}\Box_{c_1}$np' if n_k is a nn-noun (c_1 is the case of n_1). In fact, the package $\mathcal{R}$ makes it possible to apply the lock/key strategy (Moortgat 1999) which, in our case, proceeds as the following. Firstly, recursive rules K(j) are used to open each lock $\Box_{cs}$ surrounding nouns n_i ($1 \leq i < k$)[6], thus checking that they are all in construction state. Secondly, Inc(i, j) rules are used to check definiteness feature of noun n_k. Finally, the derivation can be completed by a succession of $/_0$ elimination. Therefore, it is easy to parse a compound nominal sentence such as 's_1=*ʾibnu 'l-mudarris-i fī 'lbayti*' (i.e., the man's son is in the house), the crucial steps of its topic's derivation are sketched below:

[6]The key $\Diamond_{cs}$ allows to open $\Box_{cs}$ lock since $\Diamond_{cs}\ \Box_{cs}\text{A} \vdash \text{A}$

$$\dfrac{\dfrac{}{ibnu : \Box_{cs}(\Box_{nom}np/_0\Box_{gen}np) \vdash i : \Box_{cs}(\Box_{nom}np/_0\Box_{gen}np)}\,Ax}{< ibnu : \Box_{cs}(\Box_{nom}np/_0\Box_{gen}np) >_{cs} \vdash i : \Box_{nom}np/_0\Box_{gen}np}\,\Box_{cs}E \qquad \dfrac{\dfrac{}{'lmudarrisi : \Box_{al}\Box_{gen}np \vdash m : \Box_{al}\Box_{gen}np}\,Ax}{<' lmudarrisi : \Box_{al}\Box_{gen}np >_{al} \vdash m : \Box_{gen}np}\,\Box_{cs}E$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\cdots \qquad \cdots}{(< ibnu : \Box_{cs}(\Box_{nom}np/_0\Box_{gen}np) >_{cs}, <'lmudarrisi : \Box_{al}\Box_{gen}np >_{al})^0 \vdash (im) : \Box_{nom}np}\,/_0E}{(< ibnu : \Box_{cs}(\Box_{nom}np/_0\Box_{gen}np) >_{cs}, <'lmudarrisi : \Box_{al}\Box_{gen}np >_{def})^0 \vdash (im) : \Box_{nom}np}\,Inc(def, al)}{< (ibnu : \Box_{cs}(\Box_{nom}np/_0\Box_{gen}np), 'lmudarrisi : \Box_{al}\Box_{gen}np)^0 >_{def} \vdash (im) : \Box_{nom}np}\,K(def)}{(ibnu : \Box_{cs}(\Box_{nom}np/_0\Box_{gen}np), 'lmudarrisi : \Box_{al}\Box_{gen}np)^0 \vdash (im) : \Box_{def}\Box_{nom}np}\,\Box_{def}I$$

Finally, we use higher order λ-terms[7] to represent the lexical semantics of each class of **SA** nouns. For instance, singular **SA** nouns are assigned the following meanings:

cs-noun: w_{cs} (**Rel**)	$\lambda P_{(e\to t)\to t}\ \lambda Q_{e\to t}.\ \exists$ x. $\mathbf{w}_{pred}$(x) $\wedge$ Q(x) $\wedge$ P(λy. **Rel**(x,y) $\wedge$ $\forall$ z. $\mathbf{w}_{pred}$(z) $\wedge$ Rel(z,y) $\Rightarrow$ z=x)
al-noun: w_{al}	$\lambda Q_{e\to t}.\ \exists$ x. $\mathbf{w}_{pred}$(x) $\wedge$ Q(x) $\wedge$ ($\forall$ z. $\mathbf{w}_{pred}$(z) $\Rightarrow$ z=x)
nn-noun: w_n	$\lambda Q_{e\to t}.\ \exists$ x. $\mathbf{w}_{pred}$(x) $\wedge$ Q(x)

Note that the semantics of each noun w_x is based upon a predicate $\mathbf{w}_{pred}$ representing a set of individuals that share a specific property (e.g., **teacher**, **son**...). Moreover, the meaning of cs-nouns closely depends on a relation **Rel** that binds these individuals to their annexed objects (e.g., Rel can be either a relation of possession, family-ship ...). Lastly, it is worth noticing that both cs-nouns and al-nouns meanings require uniqueness conditions. For instance, the semantics of a cs-noun w_{cs} stipulates that the intersection between the set of individuals verifying the property $\mathbf{w}_{pred}$ and the range of entities connected to the annexed object by the relation **Rel** is nothing else but a singleton.
Using the previous lexical semantics, one can easily check that the meaning of sentence s_1 is represented by the following first-order formula:

$\exists$x. **son**(x) $\wedge$ **in**(x, ι **house**) $\wedge$ $\exists$y. **teacher**(y) $\wedge$ **family-ship**(x,y) $\wedge$
($\forall$ z. **son**(z) $\wedge$ **family-ship**(z,y) $\Rightarrow$ z=x) $\wedge$ ($\forall$z. **man**(z) $\Rightarrow$ z=x)

Our study can also be applied to account for nouns built using annexation in Hebrew. In fact, this phenomenon is managed by the same range of syntactic principles in both Hebrew and **SA** as shows the following example quoted from (Wintner 2000):

(9) *yaldei* *mnahhel* *taxnot* *ha-rakkebt*
children-[cs] manager-[cs] stations-[cs] the train
'the train stations manager's children'

Adjectives

We distinguish two classes of adjectives in **SA**, namely *attributive* adjectives and *predicative* ones. Attributive adjectives are used to modify definite and indefinite nouns, they can be involved in the construction of enhanced topics. These adjectives agree with the head they modify on number, gender, case and definiteness. However, predicative adjectives are used as comments within nominal sentences; they are always indefinite and they agree with their

[7]To preserve the form-meaning correspondence, we will assume that the syntactic type np is lifted to $(s/_0\ np)\backslash_0 s$

topic on gender and number. We present in the following some examples of well-formed and ill-formed **SA** clauses involving the use of adjectives:

(10) *ʾibnu* *'l-mudarris-i* *'l-ǧamīl-u*
son-[nom] the teacher-[gen] beautiful-[nom]
'The beautiful teacher's son'

(11) **ʾibn-u* *ǧamīl-u* *'l-mudarris-i*
son-[nom] beautiful-[nom] the teacher-[gen]

(12) *'l-mudarris-u* *'l-ǧamīl-u* *'l-ʿaynayn-i/*lawn-u* *'l-ʿaynayn-i*
the teacher-[gen] beautiful-[nom] the eyes-[gen]/*color-[nom] the eyes
'the teacher with beautiful eyes'

(13) *'l-mudarris-u* *ǧamīl-u* *'l-ʿaynayn-i*
the teacher-[gen] beautiful-[nom] the eyes-[gen]
'the teacher has beautiful eyes'

Unlike some languages where the word-order between adjectives and their head is relatively free (e.g., French), **SA** attributive adjectives are *post-modifiers*, they always occur after the noun phrase they modify. Moreover, when the noun phrase is built using annexation, the adjective should be placed at the end of the whole construction (cf. ex. 10) since nouns in construct state cannot be modified (cf. ex. 11). On the other hand, we are able to build enhanced adjectival phrases thanks to annexation phenomenon (cf. ex. 12&13). In **SA**, an adjectival phrase has two constituents. The first one (i.e., the head) is either a cs-adjective (adjective in construct state) or an al-adjective (simple definite adjective) whereas the second one (i.e., the complement) is nothing else but a genitive al-noun. Hence, it is forbidden to build an adjectival phrase by combining an adjective with a compound noun phrase (cf. ex. 12).
The following table recapitulates the various syntactic constraints which manage the use of **SA** adjectives. For the sake of legibility, we only focus on definiteness and case agreements between adjectives and their heads; gender and number agreements can be added in a straightforward fashion.

SA adjectives	al-adjectives	nn-adjectives	cs-adjectives
Predicative	×	$P_2 = \Box_{def}\Box_{nom}np\backslash_c s$	$P_3 = P_2/_0\Box_{al}\Box_{gen}np$
Attributive	$A_1 = \Box_{def}\Box_{case}np\backslash_0\Box_{def}\Box_{case}np$ $A'_1 = A_1/_0\Box_{al}\Box_{gen}np$	$A_2 = \Box_{ind}\Box_{case}np\backslash_0\Box_{ind}\Box_{case}np$	$A_3 = A_2/_0\Box_{al}\Box_{gen}np$

Thanks to this type assignment, we are able to handle the different uses of **SA** adjectives in a rigorous manner. Lexical ambiguity is necessary to account for the distinct syntactic behaviors of some adjectives. For instance, al-adjectives can directly combine with their definite nouns (e.g., ' *l-ǧamīl-u* in ex. 10); in that case we use the syntactic type A_1. Otherwise, these adjectives can initially form an adjectival phrase by combining with a genitive al-noun to their right and then merge with their head to the left (e.g., ' *l-ǧamīl-u* in ex. 12). This latter case is dealt with by means of the second syntactic type A'_1.

The semantics of **SA** adjectives resembles the semantics of English adjectives at a great extent. The interested reader can find more details about this field in (Chierchia 2000).

Negation

In contrast with some languages where negation is a slightly complex phenomenon (e.g., negation in French involves a discontinuous constituent ne ... pas), nominal sentences negation in **SA** is obtained in a straightforward fashion. This is done by adding the particle *mā* at the start of the sentence as illustrated below:

(14) *mā al-māl*ʿ*-u manb*ʿ*-u 'lassa*ʿ*ādat-i*
not money-[nom] source-[nom] the happiness-[gen]
'Money is not the source of happiness'

Consequently, we can easily deal with nominal sentences negation in **MMCG** by assigning the syntactic type $\mathbf{s}/_0\ \mathbf{s}$ to the particle *mā*.
The negation of **SA** nominal sentences can also be obtained by using some *external governors* such as *laysa* (Blachère 1994; Arrajihi 1975). This latter particle also precedes the nominal sentence but it changes the default case of its comment as shows the following example:

(15) *laysa al-māl-u manba*ʿ*-a* ʾ*lssa*ʿ*ādat-i*
not money-[nom] source-[acc] the happiness-[gen]

Since the external governor *laysa* does not subcategorize for a whole nominal sentence, we will not be concerned with its formal study in this work.

Conclusion & Prospects

In this paper, we presented the first steps towards the syntactic and semantic analysis of **SA** using **MMCG** formalism. In particular, we showed that we can deal with simple and compound nominal sentence constructions using appropriate structural modules. Moreover, the meaning of these sentences is obtained in a compositional fashion thanks to the use of simply typed λ-calculus and Curry-Howard correspondence. The complete study also includes a logical treatment of the asymmetry between the different forms of verbal phrases in **SA**. This latter work will be described in a forthcoming paper.
Our ultimate goal is to build a compact **MMCG** grammar which handles a decent fragment of **SA** containing at least the frequent linguistic phenomena, e.g., ellipsis, coordination, wh-questions and anaphora. Moreover, we intend to extend our work and deal with some Arabic dialects used in different *speech communities* (e.g., Moroccan and Egyptian). We also plan to compare Arabic phenomena with other Semitic languages, notably Hebrew.
We hope that the present survey will constitute the pillar of the formal study of cross-linguistic variation between the various forms of Arabic Language.

Acknowledgements

I would like to thank my father Abdelhamid Anoun as well as my colleagues Youssef Tahir and Dina El kassas for fruitful discussions about **SA** linguistics. I am also very grateful to the anonymous ESSLLI reviewers for their useful comments and constructive suggestions. Finally, I am deeply indebted to my parents, my sister Soumaya and my fiancé Réda for their psychological support and everlasting care. Of course, all remaining errors and omissions are mine.

Bibliography

Abdel-Hafiz, A.: *Verb Agreement in SA: An analysis in the Minimality Program* Journal of Language and Linguistics, Vol 4, (2005)

Arrajihi, A.: *alttaṭbīqu 'lnaḥwī* Dar *alnnahḍat alʿarabiyat* eds, (1975)

Blachère, R.: *Eléments de l'Arabe classique* Maisonneuve Et Larose eds, ISBN 2706811293, (1994)

Chierchia, G., McConnell-Ginet, S.: *Meaning and Grammar, An Introduction to Semantics* The MIT press, (2000)

Haddad, B., Yaseen, M.: *A Compositional Approach Towards Semantic Representation and Construction of Arabic* LACL, (2005)

Heylen, D.: *Types and Sorts Resource Logic for Feature Checking* PhD thesis, Utrecht,1999.

Kreemers, J.: *The Arabic noun phrase: a minimalist approach* PhD thesis, Nijmegen University, (2003)

Moortgat, M.: *Categorial Type Logic* Chapter 2 in Van Benthem & ter Meulen (eds), Handbook of Logic and Language, (1997)

Moortgat, M.: *Meaningful patterns* In Gerbrandy, Marx, de Rijke and Venema (eds), Amsterdam, (1999)

Oehrle, R.T.: *Multi-Modal Type Logical Grammar* Borsley, R. & Borjars, K. (eds), Non transformational Syntax, Blackwell, (2001)

Ryding, K.: *A Reference Grammar of Modern Standard Arabic* Cambridge University Press, (2005)

Sibawayh: *ʿālamu 'l-kitābi* Bayrout edition, (1983)

Vermaat, W.: *The Logic of Variation, A cross-linguistic account of wh-question formation* PhD thesis, Utrecht, (2005)

Wintner, S.: *Definiteness in the Hebrew Noun Phrase* Journal of Linguistics, Vol. 36, (2000)