The categorial annotation of coordination in Arabic

Ismail Biskri & Boucif Amar Bensaber

Département de Mathématiques et Informatique – Université du Québec à Trois-Rivières
CP 500, Trois-Rivières, (QC) G9A 5H7, Canada
{ismail.biskri ; bensaber}@uqtr.ca

Abstract
The model of categorial grammars is a sufficiently flexible model to give an account of several languages and several forms of organisation of the linguistic units of these languages (coordination, subordination, etc). However, the research tasks were limited almost exclusively to European languages. A language like Arabic, in spite of the richness of the tradition of its schools of thought, which go back for several centuries, remains unexplored by this current. In our article, we show how the model of Applicative Combinatory Categorial Grammar can give an account of categorial annotation of Arabic, in particular of certain forms of Arabic coordination.

1. Coordination
Let us consider the following statements:
(i) thaâhada ettarafani bi eliithirafi elmouthabadili wa bi elimthinaii ân ayi thaharoukathin âskariyathin
(ii) youkhatibou elibnou abahou bi ihthiramin wa oumahou bi hananin
(iii) yaakoulou elqirdou elmawza wa eddoubou elâssala
(iv) elibnou youkhatibou wa abahou bi ihthiramin wa oumahou bi hananin

The statement (i) represents the coordination of two members of different nature but fulfill the same function. The statements (ii) and (iii) represent two different cases of coordination with ellipse. In both cases the verb is elided and two non-constituents are coordinated. In the first case (ii) the non-constituents [abahou bi ihthiramin] and [oumahou bi hananin] are sequences of an object (abahou and oumahou) followed by a backward modifier of an intransitive verb (bi ihthiramin and bi hananin). In the case of (iii) the non-constituents [elqirdou elmawza] and [eddoubou elâssala] are sequences of a subject followed by an object. The statement (iv) represents the case of a correlative coordination where the conjunction appears in front of each member of coordination. Correlative coordination does not seem very current in Arabic, but, nevertheless, it is not agrammatical.

Coordination is an explicit or an implicit relation which links elements of the same sort: sentences, or linguistic units which have the same function. It is expressed explicitly by markers (of coordination) which are in fact conjunctions, such as ٍّ wa (and) (In this paper, we will focus entirely on this marker).

In the scientific literature, several formalization of coordination was proposed. The syntagmatic models appeared more promising based on the arguments advanced by Pollard, Sag, and Beavers for the process of coordination with ellipses using an HPSG Grammar (Beavers, Sag, 2004) (Sag, 2003), they do not seem to yield general rules for treating all types of coordination. This is why the TCCG was proposed for treating English coordination (Beavers, 2004). The challenge facing TCCG is to use certain Categorial Grammars rules while keeping the HPSG grammar base. The model of Categorial Grammars in its various versions offered interesting perspectives but was criticized for “over-generation” mainly because of type raising rules and its incapacity to account for certain constructions in which the elements coordinated are not of comparable nature or do not fulfill the same function (Sag, 2003), though we proved in (Biskri, Desclés, 2006) that a categorial model is able to analyze a coordination of different categories.

Very few studies were devoted to the correlative coordination. Some work is, however, listed for English with the generative current (Sag, 2005; Schwarz, 1999; Hendriks, 2004) or from the point of view of translinguistic generalization (Johannessen, 2005; Zoerner, 1999; De Vries, 2005). We find two diagrams for the analysis of the correlative coordination: (a) A symmetrical analysis (Sag, 2003) for which each member of coordination is introduced by the same conjunction; (b) An asymmetrical analysis (Johanessen, 2005; Hendriks, 2004) which consists in regarding the initial conjunction as an adverbial modifier.
2. Categorial analysis

The model of Applicative and Combinatory Categorial Grammar (ACCG) (Biskri, Desclés, 1997) as most of categorial models (Dowty, 2000; Morrill, 1994; Moorgat, 1997; Steedman, 2000; Baldridge, Kruifjff, 2003), assigns syntactical categories to each linguistic unit in order to express its function. The basic categories N and S are assigned respectively to noun phrases and sentences. The orientated categories, developed from basic types by means of the two operators of types construction “/” and “\”, are assigned to the linguistic units which function like operators. For example, the category (S/N)/N is assigned to transitive verbs which are seen as operators with two operands, the first being the object of type N positioned to its right whereas the second being the subject of type N positioned to its left. In our paper, a linguistic unit u with the type X will be noted by [X : u].

According to the postulate that the representation of the language is with three levels (Desclés, 1996) (Desclés, 1990) (Shaumyan, 1998): (i) the morphs-syntactical level; (ii) the predicative level; (iii) The cognitive level, the ACCG makes it possible, by means of rules, to give an account of: (1) to ascertain syntactic correctness; (2) to progressively construct the semantic functional interpretation; (3) to allow a functional analysis of a linguistic marker (example: wa (and),...).

The premise of each rule is a concatenation of linguistic units with oriented types. The consequence of each rule is an applicative typed expression with possible introduction of one combinator. The type-raising of one unit u introduces the combinator C*; the composition of two concatenated units introduces the combinator B.

Combinators are used in order to construct semantic interpretation. Each combinator is introduced or eliminated by a β-reduction. For illustration, we present the β-reduction rules of combinators used in this paper Φ, B and C* (U1, U2, U3, U4 being typed applicative expressions which function either like operators or like operands):

Let us deal now with a simple example in Arabic:

\[\text{تدعم الحرية الديمقراطية} \] (that we will transcribe for the needs of our paper by means of Latin characters: Thoudaimou elhouriathou eddimouqratiyatha (Freedom reinforces democracy))

1. [(S/N)/N : thoudaimou]-[N : elhouriathou]-[N : eddimouqratiyatha]
2. [(S/N)/N : thoudaimou]-(S/S/N)[C* elhouriathou]-[N : eddimouqratiyatha]
3. [S : (B : C* elhouriathou) thoudaimou] [N : eddimouqratiyatha]
4. [S : (B : C* elhouriathou) thoudaimou eddimouqratiyatha]
5. ((B : C* elhouriathou) thoudaimou) eddimouqratiyatha)
6. ((C* elhouriathou) (thoudaimou eddimouqratiyatha))
7. (thoudaimou eddimouqratiyatha) elhouriathou)
8. thoudaimou eddimouqratiyatha elhouriathou

The transitive verb in Arabic is a predicate which needs its two arguments (the subject and the object) on its left (its right with the Latin transcription). Contrary to French and English, Arabic being VSO, the subject is intercalated between the verb and the object. In the traditional Arabic grammar, it is easy to differentiate the subject from the object being given the presence in the sentence of morphological indices. Concretely, the categorial syntactical type (S/N)/N is assigned to the transitive verb thoudaimou. In other words the verb, here, is seen as an operator whose first operand is the object with type N positioned to its right in order to construct a complex operator whose operand would be the subject of type N positioned to the right. The assignment of the types is made at step 1. Steps 2 to 4 occur in the morphs-syntactical level. They consist in applying categorial rules. The type raising rule (<T) is applied to elhouriathou at step 2 in order to obtain the operator (C* elhouriathou) whose operand would be the result of the application of the transitive verb to the object. In Arabic analysis, the strategy of application of categorial rules should not be the same as in French or English. For recall, this one consists of the application of type raising rules only if the other rules are not
possible. That would make in the case of Arabic the subject as the first operand of the transitive verb, what is against the tradition established in linguistics. The sentence is seen as syntactically correct at step 4, because we get the type S. At step 5, a natural deductive process of combinatory logic is carried out in the predicative level. At step 8 the functional semantic interpretation is constructed. It expresses the effective order operator/operator of the linguistic units of the initial statement.

The example shown here is a sentence known as verbal being given that it begins with the verb. In Arabic, the majority of the sentences start with the verb, however it is not excluded to have sentences known as nominal

To avoid any form of over-generation, within the ACCG model we provide meta-rules to control the application of the type raising. Thus, in first case (respectively the second) the type raising rule ($<$T) (respectively ($>$Tx)) is started by the meta-rule M1 (respectively M2).

M1: If $u_1$ is of type $(Y/N)/Z$ and $u_2$ of type $N$, then we apply the type raising rule ($<$T) to $u_2$:

\[
(N : u_1 \implies Y(Y/N) : (C^* u_1))
\]

M2: If $u_1$ is of type $N$ and $u_2$ of type $(Y/N)/Z$, then we apply the type raising rule ($>$Tx) to $u_1$:

\[
(N : u_1 \implies Y(Y/N) : (C^* u_1))
\]

With the model of the ACCG, we consider that the conjunction applies to two linguistic units fulfilling the same function and that the linguistic unit which results from coordination also inherits this function. This postulate takes form in Lambek's scheme of type $(X/X)/X$ associated to the conjunction $\land wa$ ($\text{and}$). In this type the variable $X$ can be unified with any basic or non-basic type.

Let us show the categorial analysis of the sentence (ii):

\[
1. (\text{N : elhouriyathou} \land \text{thoudaiimou}) \land \text{eddimouqratiyatha}
\]

\[
2. ([S/N] : elhouriyathou) \land \text{eddimouqratiyatha}
\]

\[
3. ([S/N] : (B \text{ elhouriyathou} \land \text{thoudaiimou}) \land \text{eddimouqratiyatha})
\]

\[
4. ([S/(S/N) : (B \text{ elhouriyathou} \land \text{thoudaiimou}) \land \text{eddimouqratiyatha}])
\]

\[
5. ([B \text{ elhouriyathou} \land \text{eddimouqratiyatha})]
\]

\[
6. ((C^* \text{ elhouriyathou} \land \text{thoudaiimou} \land \text{eddimouqratiyatha})]
\]

\[
7. \text{thoudaiimou eddimouqratiyatha elhouriyathou}
\]

\[
8. \text{thoudaiimou eddimouqratiyatha elhouriyathou}
\]

where the subject is placed before the verb : (elhouriyathou thoudaiimou eddimouqratiyatha)

الحربية تدعم الديمقراطية However, the traditional grammar of Arabic considers the subject as being the "moubhada" and the verbal syntagm, formed by the transitive verb followed by an "erased" pronoun clitic and an object, as the "khabar". It is a form of topicalization of the subject in which the pronoun clitic which would have to take the place of the subject (to make a parallel with the French topicalization) on the right of the verb is erased. The application of the type raising rule ($>$Tx) in the analysis which follows takes in account this fact.
The analysis starts with the assignment of the syntactic categories to the lexemes. For recall, each syntactic category describes the way in which a lexeme operates on its arguments. The category \( (X \times X) / X \) assigned to the conjunction is in fact a scheme of type which describes the conjunction like an operator whose first and second operands, of type \( X \), are respectively the second member and the first member of the coordination. The type of the coordination \( (S / N) / (S / N) / N \) which will be substituted to \( X \) is known after the construction of the second member of coordination (step 9). Steps 1 to 15 represent the application of ACCG rules. With these steps we verify the correctness of the sentence (the type \( S \) obtained at 15). A first structural reorganization is applied at steps 5 in order to extract the operand of bi-ilthiramin. A second structural reorganization is applied at steps 10 and 11 in order to extract the first member of the coordination. The structural reorganization (Biskri, Desclés, 1997) is based mainly on the reduction and/or the introduction of some combinators into a combinatory expression to give an equivalent but differently structured combinatory expression. At the step 2, the type raising rule \(<T>\) is controlled by the meta-rule M1. But at step 8 the same rule is controlled by the following meta-rule:

7. \[ S / [(B (C* elqirdou) yaakoulou) / elmawza] - [(XX) / X : wa] - [S / (S / N)] / [N : [C* eddoubou] (C* elâssala)] \(<T>\)
8. \[ S / [(B (C* elqirdou) / C* elmawza) / yaakoulou] / [(XX) / X : wa] - [S / (S / N)] / [B (C* eddoubou) (C* elâssala)] \(<T>\)
10. \[ (S / N) / N : yaakoulou] - [S / (S / N)] / [B (C* elqirdou) / C* elmawza] - [(S / (S / N)) / (S / (S / N))] / [S / (S / (S / N))] / [B (C* eddoubou) (C* elâssala)] \(<T>\)
11. \[ (S / N) / N : yaakoulou] - [S / (S / N)] / [B (C* elqirdou) / C* elmawza] - [(S / (S / N)) / (S / (S / N))] / [S / (S / (S / N))] / [S / (S / (S / N))] / [B (C* eddoubou) (C* elâssala)] \(<T>\)
12. \[ S / [(S / (S / N))] \(<T>\)
13. \[ ((wa [(B (C* eddoubou) (C* elâssala)]) / (B (C* elqirdou) / C* elmawza)] / yaakoulou) \(<T>\)
14. \[ ((Φ ∧ (B (C* eddoubou) (C* elâssala))) / (B (C* elqirdou) / C* elmawza)] / yaakoulou) \(<T>\)
15. …
16. \( (Φ ∧ (B (S / (S / N))] / (S / (S / (S / N))] / [S / (S / (S / N))] / [S / (S / (S / N))] / [B (C* eddoubou) (C* elâssala)] / yaakoulou)

It is interesting to notice that, in the analysis of this statement, the construction of the second member of coordination requires the application, at steps 5 and 6, respectively, of two crossed type raising rules (forward then backward) \( (>T_x)\) and \( (<T_x)\). The application of these rules is done thanks to meta-rules M4 and M5.

M4 : For a concatenated sequence u1-u2-u3, if u1 is the conjunction \( 3 \) wa, u2 is of type S and u3 of type \( Z \), then we apply the type raising rule \( (>T_x)\) to u2 : \[ N : u2 ==> S / (S / N) / (C* u2) \]

M5 : For a concatenated sequence u1-u2-u3, if u1 is the conjunction \( 3 \) wa, u2 is of type S / (S / N) and u3 of type \( Z \), then we apply the type raising \( (<T_x)\) to u3 : \[ N : u3 ==> (S / (S / N)) / (S / (S / N)) / (C* u3) \]

The second member of the coordination is a complex operator (whose operand is the transitive verb) which in the case of our example can be only one combination of the type raised subject with the type raised object. The choice of each type raising is determined by the position of the operand in the linear structure. Thus, \( (>T_x)\) applied to the object is justified by the fact that the verb which is the operand of \( C* elâssala \) precedes the object in the linear structure, without, however, to be contiguous for it. \( (>T_x)\) applied to the subject is justified by the fact that its operand would be in the linear structure a combination of the verb and the type raised object, this latter being positioned on the right of the subject. wa, in this case too, is a linguistic operator.
The first member of coordination is delimited by the first conjunction \(\text{wa1} \) (initial one) and the second conjunction (coordinating one) \(\text{wa2}\). A consequence of that: the category of the coordination is obtained after the construction at step 5 of the first member of coordination \(\text{Bi ihthiramin (C* Abahou)}\) with the type \(\text{S(N/N)}\). The category \(\text{X/(X/X)}\) is assigned to the initial conjunction \(\text{wa1}\) in order to express its role of operator whose first operand is the first member \(\text{Abahou bi ihthiramin}\) of the coordination and the second operand is the result of the application of the coordinating conjunction \(\text{wa2}\) to the second member \(\text{Oumahou bi hananin}\). This category is in agreement with the fact that the presence of an initial conjunction is strongly dependent on the presence of a non-initial conjunction. However, it makes it possible to classify our analysis neither in the class of the symmetrical analysis nor in the class of the asymmetrical analysis as described in the literature. Indeed, on the one hand this category is different from the category assigned with the non-initial conjunction and thus defines the conjunction initial as non-coordinating, and on the other hand it is not compatible with the category of an adverb. The second conjunction \(\text{wa2}\) which is coordinating still have the scheme of type \(\text{(XX)/X}\). Steps 14 to 24 reduce combinators in order to construct the functional semantic interpretation: \((\wedge \text{(bihanin\ (youkhatibou Oumahou)) \ (bi ihthiramin (youkhatibou Abahou)) Elibnou})\), which is structured like a conjunctive clause too. This functional semantic interpretation is exactly the same as the one produced through the analysis of elibnou youkhatibou abahou bi ihthiramin wa oumahou bi hananin. At the step 17 the linguistic predicate \(\text{wa1}\) (and) is replaced by its meaning \(\text{C*}\) in the cognitive level. \(\text{wa1}\) is not a coordinating conjunction. It is just an operator who “types raise” the first member of coordination to produce a complex operator \((\text{wa1} \text{ (Bi ihthiramin (C* Abahou)) Elibnou})\) whose operand is the result \((\text{wa2} \text{ (bihanin (C*Oumahou)) youkhatibou Elibnou})\) of the application of the coordinating conjunction \(\text{wa2}\) to the second member of coordination \(\text{B bihanin (C*Oumahou)}\). At the step 19 the linguistic predicate \(\text{wa2}\), because of its distributive and conjunctive nature, is replaced by its meaning \(\Phi \wedge\) in the cognitive level. In short, the initial conjunction does not function like a simple conjunction. It is different by its not coordinating nature. It is mainly useful to reinforce coordination,
either by eliminating ambiguity as in French (Biskri, Rochette, 2007) or by determining the first member of coordination as well as the type of coordination. The analysis of correlative coordination is asymmetric, even if the initial conjunction does not function like an adverb.

3. Conclusion

With this work, it arises several fundamental results. From a theoretical point of view, with the ACCG, thanks to the use of the combinatory logic, and the meta-rules we prove, on concrete examples, that it is a very solid formalism and very flexible device for the analysis of coordination in Arabic. The analysis carried out confirms that the categorial operators and the combinator of the combinatory logic construct, starting from the concatenated structure, the right interpretations of the statements; these interpretations are expressed in functional terms (operator/operand). In addition, these results show that in spite of criticisms (even if all are not sufficiently verified) concerning all categorial models, it remains that the categorial approach is enough flexible to be adapted to Arabic, only meta-rules should be exclusive to each language.

On another hand, we prove, in a general way, that Arabic language is not just a linear succession of linguistic units. Certain units function like complex operators. It is useful to formalise their meaning. The model of representation of the language with three levels (morphs-syntactical level/predicative level/cognitive) is in adequacy with this reality.

References


