

A Formal Description of Arabic Syntax in Definite Clause Grammar

Hisham El-Shishiny

IBM Scientific Center, 56, Gameaat El-Doual El-Arabeya St.

Mohandesseen, Cairo, Egypt

Abstract

Arabic has some special syntax features which lead to complex syntax structures. We have developed a formal description of Arabic syntax in Definite Clause Grammar. This grammar is characterized by its high descriptive power due to its dual formulation in terms of functions and in terms of grammatical categories. The developed grammar has a high coverage of Arabic language and has context sensitive capabilities. It is suitable for the advanced applications of natural language processing.

1. Introduction

Arabic has some features which lead to complex syntax different from those of the European languages. Moreover, Arabic grammar exists only in a descriptive form and there is no comprehensive formal representation for it so far.

The few attempts to give a formal model for Arabic sentences were based on transformational generative grammar – [1], [2], [3] and others – but some linguists adopted more recent linguistic models such as Lexical Functional Grammar [4], dependency grammar [5] and functional grammar [6].

Selecting a suitable grammar formalism for Arabic and the formulation of the grammar itself is the subject of a big debate in the Arab countries nowadays. Our description of Arabic syntax in Definite Clause Grammar is different from the attempts undertaken so far.

2. Definite clause grammar for Arabic syntax

In this grammar, the arguments of non

terminals are used to hold special features of Arabic words such as the definitely or indefinitely determined feature of nouns which is essential to recognize many structures. Also, the ability of Definite Clause Grammar to provide a general treatment of context sensitivity through the proper use of arguments of non terminals makes it easy to account for agreements such as number, gender and person agreements.

Another important characteristic of Definite Clause Grammar, which suits Arabic, is the possibility of imposing extra conditions on the constituents of a phrase which must be satisfied for a rule to be valid [7].

3. The developed formal grammar

Due to the fact that there is no single basic word order for Arabic sentences, three basic sentence types were defined:

- a. Nominal sentence: a sentence that does not contain a verb or contains a verb which follows the subject.
- b. Verbal sentence: a sentence that contains a verb which precedes the subject.
- c. Sentences with special structures such as vocative sentences.

Sentences are further classified functionally and according to their modalities.

Larger sentences are also described in this grammar by conjoining sentences or embedding simpler sentences.

We have used the grammatical categories together with the functional roles to define the syntactic structures. Grammatical categories alone were considered not sufficient to describe all the structures

since word categories in Arabic are globally classified into verbs, nominals and particles only. Nouns, adjectives and adverbs for example are differentiated in the sentence according to their functions.

This dual formulation increases the descriptive power of the grammar. In order to use this dual formulation to define the syntactic structures, some non standard definitions of grammatical categories were used.

A nominal sentence (ns) is defined formally as composed from a 'mobtadaa' phrase, mbp, and a predicate phrase, predp:

$$\text{ns}(\text{ns}(\text{MBP}, \text{PREDP})) \rightarrow \text{mbp}(\text{N}, \text{G}, \text{m}, \text{-}, \text{MBP}), \text{predp}(\text{N}, \text{G}, \text{n}, \text{PREDP}).$$

A mbp is defined as a noun phrase (np) that can have 'motaalkat' which are either a prepositional phrase (PP) or adverbial.

A predp can be one of the following:

- a noun phrase that can have 'motaalkat'
- a prepositional phrase or adverbial
- a sentence which can be :
 - a verbal sentence
 - a nominal sentence (under certain conditions).

Verbal sentences (vs) can be followed by either a subject phrase, sp, or a subject phrase and a complement phrase, cop:

$$\begin{aligned} \text{vs}(\text{vs}(\text{VP})) &\rightarrow \text{vp}(\text{T}, \text{V}, \text{S}, \text{P}, \text{VP}). \\ \text{vs}(\text{vs}(\text{VP}, \text{SP}, \text{COP})) &\rightarrow \text{vp}(\text{T}, \text{a}, \text{S}, \text{P}, \text{VP}), \\ \text{sp}(\text{-}, \text{-}, \text{S}, \text{-}, \text{P}, \text{SP}), \text{cop}(\text{M}, \text{COP}), \{\text{M}=\langle \text{T} \rangle\} \\ \text{cop}(\text{o}, \text{cop}(\text{I})) &\rightarrow \text{I}. \end{aligned}$$

For verbs in the passive voice the sp is replaced by 'naabfael' phrase (nsp), which can be either a np or pp.

For non-terminal argument definitions the reader is referred to sections 3.1.1., 3.1.2 and 3.1.3. The values of the arguments indicated above are: a for active voice, n for indefinite determination, m for definite determination and - for acceptance of any value.

Verb phrases (vp) are defined as follows:

$$\text{vp} \rightarrow (\text{cop}), (\text{particle}), \text{verb}, (\text{cop}).$$

whereas noun phrases (np) are defined as follows:

$$\text{np} \rightarrow (\text{particle}), \text{noun}, (\text{post-modifiers}).$$

Subject phrases are noun phrases whereas complement phrases modify the verbs and are one or more noun phrases (such as the direct object) and/or prepositional phrases.

The cop category is used to cater for the cases where the sp separates the verb from some of its modifiers, which is a characteristic of Arabic syntax.

Sentences with special structures have different structures which vary from one case to another.

Sp and cop can be relative phrases and in some cases the predp can precede the mbp. Categories with no equivalent in English are indicated between single quotes and categories inside the brackets are optional.

3.1. Non Terminal arguments

Non-terminal arguments are used to hold features of Arabic words necessary for recognizing some structures and for allowing agreements between the different constituents of the sentence. Arguments are introduced in the lexical entries of words and are inherited by the phrase in which the words are constituents. In this grammar, arguments are defined as follows:

3.1.1. Arguments associated with verbs

T transitivity, V voice, P person, S semantic feature.

3.1.2. Arguments associated with nouns

N number, G gender, D definite or indefinite determination, P person, S semantic feature.

3.1.3. Arguments associated with pronouns

N number, G gender, P person.

M is an argument associated with some non-terminals to test the possibility of modifying a verb with objects.

3.2. Use of semantic features to reduce ambiguity in Arabic sentences.

Semantic features are used with verbs

and nouns in order to test the subject verb semantic agreement.

The use of the described semantic features in the developed grammar helps in the differentiation between the subject and the object in Arabic sentences, since it is not possible on purely grammatical basis (except in some exceptional cases) to differentiate between them.

3.3. Examples

In this section, examples of the syntactic structures of a nominal sentence and a verbal sentence according to the developed grammar are given.

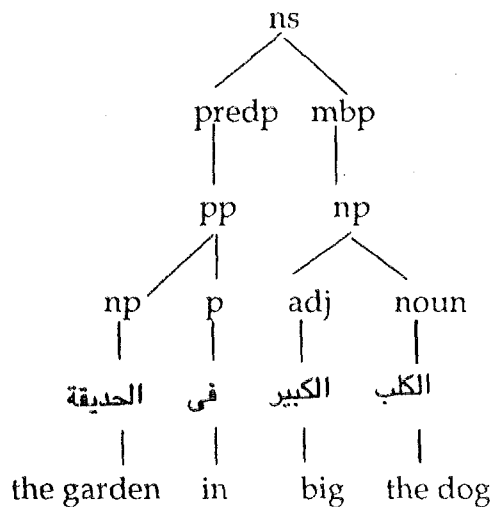
3.3.1. Example of a nominal sentence:

الكلب الكبير في الحديقة

Which means:

the big dog is in the garden.

In arabic, the definite article is a prefix, there is no copulative verb and direction of writing is from right to left.

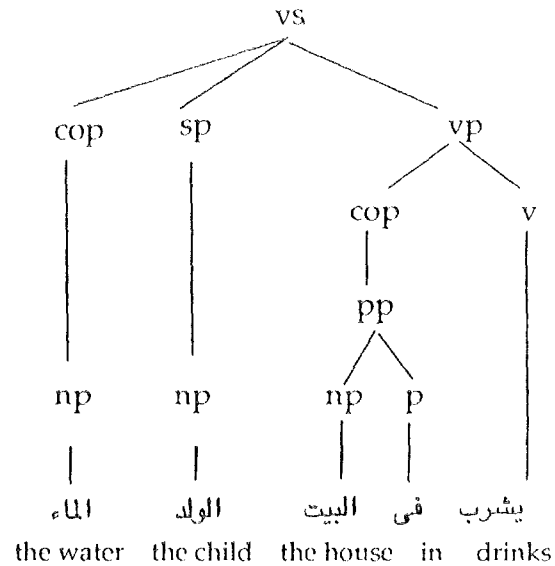


3.3.2. Example of a verbal sentence:

يشرب في البيت الولد الماء

Which means:

the boy drinks the water in the house



4. Conclusion

This grammar has been implemented in a syntactic analyzer [8] developed in Prolog on a PC/XT-286 and tested on a large number of Arabic sentences.

It is hoped that the developed grammar will contribute to the efforts undertaken recently in the Arab countries towards the development of a comprehensive, agreed upon, formal grammar for Arabic syntax.

References

- [1] Bakir M. 'Aspects of clause structure in Arabic: a study of word order variation in literacy Arabic', Ph.d., Indiana University, 1980.
- [2] Al-Khuli M. 'A contrastive transformational grammar: Arabic and English', Leiden: Brill, 1979.
- [3] Ayoub G. 'Structure de la phrase verbale en Arabe standard', Ph.D., in: Analysis/Theories, 1981.
- [4] Fehri F. 'Complementation et amophore en Arabe moderne: une approche lexicale fonctionnelle', These de Doctorat d'Etat, Universite de Paris 3, 1981.
- [5] Owens J. 'Structure, Class and Dependency: Modern linguistic theory and the Arabic grammatical tradition', in: Lingua 64, 1984.
- [6] Moutouakil A. 'Pragmatic functions in a functional grammar of Arabic', Dordrecht: Foris Publications, 1989.
- [7] Pereira F., Warren D. 'Definite clause grammar for language analysis- A survey of the formalism and a comparison with transition networks', Artificial Intelligence, Vol. 13, pp. 231 - 278, 1980.
- [8] Hisham El-Shishiny. 'A syntactic analyzer for Arabic sentences', IBM-CSC Technical Report # 32, 1989.