

A PHRASE STRUCTURE GRAMMAR OF THE ARABIC LANGUAGE

AYMAN ELNAGGAR¹

ABSTRACT

A lot of work has been done in the field of natural language processing (NLP) for ARABIC. Few researchers have tried hard to apply on ARABIC the methods of computational linguistics, as for example, Definite Clause Grammar (DCG) [1] and Augmented Transition Networks (ATN) [2]. Because there is not a modern linguistic model for ARABIC grammar within the frame of computational linguistics, the results achieved are not comparable with that achieved on ENGLISH.

In this paper, we represent the phrase structures covering ARABIC. They are completely different from ENGLISH.

1. INTRODUCTION

Since ARABIC is a rich and highly inflected language [3], there are grammatical categories in ARABIC which do not exist in other foreign languages such as in ENGLISH [4]. For example, there is an ARABIC category ² called 'MAFOOL MOTLAK' which has no equivalent in ENGLISH. Moreover, there are syntactic categories in ARABIC which could themselves be a whole sentence or even a prepositional phrase. An example of such a category is the adverb. In addition, there are other differences in the order of the constituents within the sentence. For example, in ARABIC, adjectives follow the noun which they modify.

In fact, ARABIC has its own structure, which is completely

¹ 21 KASER EL-EINI St. CAIRO, EGYPT.

² When there is no equivalent translation available for any category in ENGLISH, the transliterated form will be shown between single quotes.

different from ENGLISH and hence, requires a different treatment.

2. ARABIC AND COMPUTATIONAL LINGUISTICS

As the need increases to a model of ARABIC, ARABIC linguists try to re-organize the grammatic system in a way that may help the researchers in the field of computational linguistics. Linguists is traditionally concerned with formal, general, structural models of natural language. Linguists, therefore, have tended to build formal models which allow to capture, as much as possible, the regularities of language and make the most appropriate linguistic generalizations. Little or no attention has been paid to characterize the language itself, ignoring the mechanism that produce it or decipher it.

So, the researchers spent most of the time in re-building the models given by the linguists. Their aim was to make these models acceptable in computational field (synthesis, generation and translation).

2.1 THE KEY

What is new in this work is that we classified all the noun phrase modifiers into two categories. **Category one**, noun phrase post-modifiers, in which the ordered sequence of these modifiers is significant and it is the key to construct these noun phrase structures. These modifiers are noun complement (NC), adjective (ADJ), apposition (APP) and correlation (COR). **Category two** which consists mainly from VP modifiers like adverb.

The above categorization has given us the ability to focus on each category and finally the NP structures were obtained.

3. THE NOUN PHRASE STRUCTURES

We have **Eight** structures covering all the categories of noun phrases. **FIVE** of these structures are given here.

The notation used in describing these structures is :-

() = The category inside these brackets is optional.

*() = The category inside these brackets could be repeated more than once.

[] = The category inside these brackets exists once at least.

*[] = The category inside these brackets could be repeated but it should exist once at least.

Representation is in Chomsky's Normal Form and should be interpreted from right to left .

The structures given here consist mainly of NOUN (N) as the basic unit and its **maximum allowable post-modifiers** which could follow it . These post-modifiers have a specific sequence which is the key to construct these structures and according to the number of these post-modifiers, the structures vary . The NP structures as well as some of the ARABIC categories they constitute, are given in Fig.(1) .

Example of a simple Arabic verbal sentence is given in Fig.(2) . In this example, the object of the sentence has all the post-modifiers given in the structure NP1 and It is clear that some of these modifiers have their own embedded structures shown in the embedded rectangles .

4. MODIFICATION

'MAFOOL MOTLAK' is a special modifier in Arabic which has a different forms depending on its post-modifiers. We have first [5] considered its structure as:

*(ADJ) (NC) NOUN

But later on, we have found out that both NC and ADJ can not exist simultaneously and only one of them should exist alone. So, we have split this structure to three structures

NP3, NP4 and NP5 published here.

ACKNOWLEDGMENT

I would like to thank Dr. T. ANBER⁽¹⁾ for her valuable discussion and support . Also, I would like to thank Dr. H. MAHGOUB⁽²⁾ for his comments reviewing the paper and who, first, implemented most of the structures given here.

CONCLUSION

We have tried here to present the paper in a frame work showing that the study of computational modeling of ARABIC faces many problems. It is evident that the ARABIC possesses a certain symmetry as regards its structure, which leads itself easily to computation. This work is the basis of any further computational processing of ARABIC, leading to semantic analysis and eventually machine translation .

REFERENCES

- [1] 'TWO ARABIC SYNTACTIC ANALYZERS', HISHAM EL-SHISHINY & AYMAN ELNAGGAR, PROCEEDINGS OF THE SECOND CONFERENCE ON ARABIC COMPUTATIONAL LINGUISTICS, KUWAIT , NOV., 1989 .
- [2] 'AN EXPERT SYSTEM FOR UNDERSTANDING ARABIC SENTENCES', MAHER S. AHMED, M.Sc. THESIS, CAIRO UNIVERSITY, 1988.
- [3] 'AL-NAHO AL-MOSAFFA ', Dr. MOHAMMED EID, CAIRO UNIVERSITY .
- [4] 'NATURAL LANGUAGE PARSING AND LINGUISTIC THEORIES', U. REYLE & C. ROHRER, D. REIDEL PUBLISHING COMPANY, 1988 .
- [5] ' A FINITE STATE AUTOMATA OF THE ARABIC GRAMMAR', AYMAN ELNAGGAR, PROCEEDINGS OF THE 'IEEE INTERNATIONAL WORKSHOP ON TOOLS FOR AI', GEORGE MASON UNIVERSITY, OCT., 1989.

(1) Professor of computational linguistics, AL-ALSON FACULTY, AIN-SHAMS UNIVERSITY.

(2) Project Leader, IBM Cairo Scientific Center.

CATEGORY	NP STRUCTURES
OBJECT	NP1 = (COR) (APP) (ADJ) (NC) N
ADJECTIVE	NP2 = (NC) N
'MAFOOL MOTLAK' (CASE 1)	NP3 = N
'MAFOOL MOTLAK' (CASE 2)	NP4 = [ADJ] N
CORROLATION	NP5 = [NC] N

Fig.(1) The NP structures with some of the ARABIC categories they constitute .

قابل الطالب رئيس جامعة القاهرة الجديد د.م. سلامة نفسه

ITS PROPER ENGLISH EQUIVALENT IS

THE STUDENT MET THE NEW HEAD OF CAIRO UNIVERSITY Dr. M. SALAMA HIMSELF

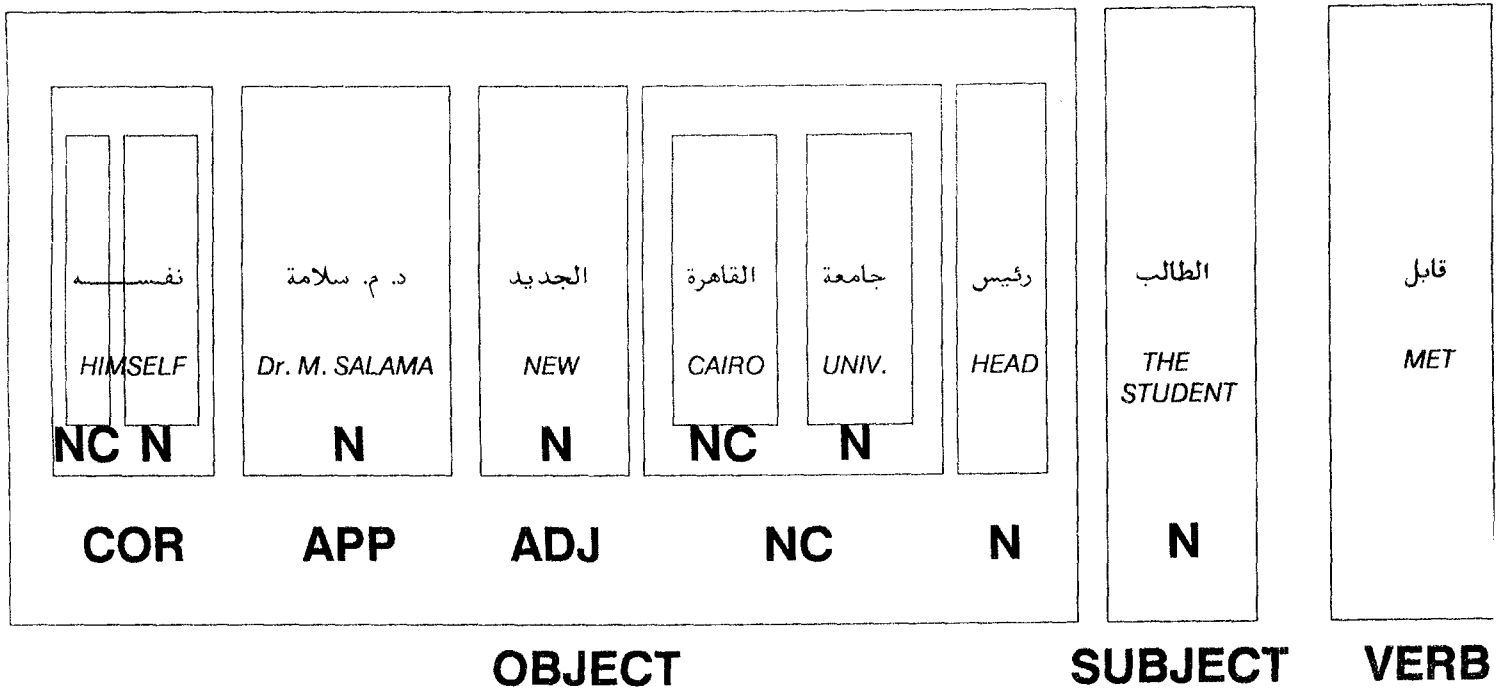


Fig.(2) Example of a simple ARABIC verbal sentence .