Lexicon Grammar within the Defining Matrix Analysis Model

Amr Helmy Ibrahim

Professeur à l'Université de Franche-Comté – UMR 8139 / LaLICC (Université de Paris-Sorbonne / CNRS)
5, rue Louis Léon Lepoutré – F-94130 Nogent-sur-Marne (France)
amr.ibrahim@free.fr

Abstract

This paper argues why the natural development of Maurice Gross's lexicon-grammar framework that has it's roots in Zellig Harris's applicative grammar based on a clear cut distinction between report and paraphrase is Amr Ibrahim's Defining Matrix Analysis Model based on different grammatically driven reductions of the redundancies of the largest defining matrix of an utterance built through a series of controlled paraphrases.

The thread that leads you through the various developments of Lexicon-Grammar, since it was set on by Maurice Gross at the end of the sixties, is that any word should get a full description of its grammatical function and, subsequently, of its contribution to the building of meaning in a discourse, according to its position and probability of occurrence within the argument structure of the simple sentence, be it overtly expressed or necessarily reconstructed, in which this word occurs.

That is why the first stage of the work achieved at Maurice Gross's LADL was mainly a thorough survey of argument structures within one language, French, in order to elaborate a typology of operators in this language so as to build a theory predicting the conditions and constraints under which these types of operators could occur in any of the world languages.

The framework of this project, on top of the general structural and transformational one, inherited three key principles that became more and more important in the late work of Zellig S. Harris.

The first, derived from his conception of applicative grammar and the function of redundancy in language states that: "Nous avons observé que tous les opérateurs, quel que soit l'ensemble auquel ils appartiennent ou leur position, et qui ont des contraintes d'arguments données, ont une transformée synonyme dans un ensemble moins contraint de ces opérateurs. Ceci revient à dire que pour toute contrainte d'arguments, il existe au moins un ensemble ou une position qui est grammaticalement moins contraite que toutes les autres et qui contient des transformées synonymes de chacun des opérateurs qui ont cette contrainte d'arguments. Nous pouvons appeler cet ensemble distingué ou cette position distinguée la source descriptive de tous les autres opérateurs, puisque nous pouvons les "dérirer" de cette source, mais nous ne pouvons pas dériver tous les membres de cette source à partir des éléments éventuellement moins nombreux et plus contraints des autres ensembles opérateurs associés à la contrainte d'arguments donnée. (...) Tous les discours sont formés par des mots opérateurs apparaissant (comme des introductions successives) dans la position distinguée (ainsi définie) et en relation avec les arguments nécessaires. Mais ce n'est qu'un sous-ensemble des discours qui est formé par ce procédé; d'autres subissent en plus (anc) processus de réduction. Dans nombre de langues, la position distinguée de l'opérateur est la même que les arguments..." (Harris 1976: 16-7).

The second states that metalanguage is in the language (Harris 1968: 17) and that this entails two types of transformations and even two systems of Grammar: Report and Paraphrase (Harris 1969). The Report referring to the system by which language describes itself.

The third states that information is a matter of argument requirement and probability of occurrence regardless of the grammatical category involved (Harris 1988).

As long as the descriptions of languages dealt with verbs the general framework of the Lexicon-Grammar looked fairly similar to that of most structural or even transformational approaches and many scholars did not pay much attention - even inside the house ! - to Harris's three principles, most of them considering them more relevant to the philosophy or epistemology of linguistics rather than to its pragmatic methodology.

Things began to change after the publication of the now famous paper "Sur quelques groupes nominaux complexes" (Gross 1976) where Maurice pointed out what is considered as one of his most genuine discoveries: the possibility of a double analysis of a simple sentence structure when its verb is a support (light) verb. During five years Maurice Gross will shift his attention towards nominalizations. He will come little by little to the conclusion that the descriptions of argument structures when revisited with the notion of support need a thorough aggiornamento (Gross 1981) and that the key principles of his master Zellig could lead to a brand new methodology in which the formal
search of the core of a predication leads to a noun dependent argument structure.

In a sentence like:

(1) Joanna caught a cold

Nobody will ever think that, from a semantic point of view, the core of the predication could be elsewhere that in cold. Moreover, if you are speaking of the present state of Joanna you would say:

(2) Joanna has a cold

that is, you would change lexically the verb giving thus evidence that it is in the noun and not in the verb that lies the heart of the matter. Besides, you could say to her:

(3) You should treat your cold using the noun cold without the help of the verbs to catch or to have. Going further you could reduce the verb to treat in /3/ to its basic noun source:

(4) My advice: a quick treatment of your cold.

A sentence more likely to be translated without many ordering and grammatical changes in almost all the languages of the world.

This means that the source argument structures in these cases are noun dependent and that the relation to be investigated is that of the actualizing process of nouns mainly through determination and support components leading to the completeness of the sentence. This also shows what is really on when the noun argument structure becomes - in some cases - a verb argument structure, as in:

(5) He had to follow a long course of treatment before recovering.
(5') He was treated for a long time before he could recover.
or in

(6) Do you think him strong enough to undergo this course of treatment.
(6') Do you think him strong enough to be treated this way.

Nevertheless, shifting to noun argument structure is not indeed a happy conclusion for everybody since it adds a bundle of difficulties to the language automatic data processing. The following 15 years showed in some cases the need for heavy local grammars dealing with appropriateness phenomena linked to the noun argument structure. Moreover, in some developments, it turned grammar to be nothing but a huge dictionary and most dictionaries to be a collection of sets.

Hence, emerged a still often debated issue regarding the nature of appropriateness within argument structures. Should we adopt a combinatory device or a pre-established collection of sets to handle such differences as between to follow a course of treatment and to undergo a course of treatment or to catch a cold and to have a cold, with its aspectual characteristics, on one hand, and on the other hand, the free from aspect differences between to hold a debate or a meeting but to form friendships or to pay a visit, to take a walk but also a decision, to do the cooking as well as a drawing? And what about similar relations of appropriateness within overt semantic discrepancies as in to make love, to make a statement or to make a trip? Without forgetting that whatever be the answer it should be the same in all languages. Therefore, if you choose a combinatory device we have to account for differences that have neither morphological nor semantic counterparts and if we choose to increase the collection of sets we have to account for the freedom of movement and the transformational properties of support (or light) verbs like: to hold, to pay, to form, to do, to take, to make, and so on.

The systematic analyses carried on within the framework of lexicon-grammar on many languages, namely roman, since more than thirty years, as well as the computational data processing linked to these analyses, give us today a panoramic view of the main conditions under which the insertion of lexicon units into syntactic structures is achieved. These structures often hold two functions: on one hand they help draw a distinction between usages, on the other hand they define the formal frame which, while ascribing the place of the lexical unit in the hierarchy of lexicon, will determine the specific processing worth to be applied in order to insert it into the discourse, that is to link it to a network of other units in order to construct a meaning.

Actually, we assume that the constraints that has been observed, do not only allow to list regularities and irregularities within the main or the secondary distributional and transformational properties, but allow to put forward verifiable hypothesis about the nature of these regularities or irregularities and, above all, about the principle that puts them together in order to construct a meaning in a discourse, provided that the tool describing regularities and irregularities is part of the described utterance or at least worded in an utterance strictly equivalent, from an informational point of view, to the described utterance. If such an equivalence is found, it means that the nature of the grammatical constraints that govern these regularities and irregularities can be described, specified and explained through redundant paraphrases. Actually, this is possible as long as these redundancies lift or at least reduce significantly the constraints observed in the source utterances.

For instance, when I give as equivalent to the source utterance:

(7) I travelled for five hours since I left Krakow

the redundant utterance:

(7') The occurrence of my travelling from the moment of my leaving of Krakow has a duration of five hours.

I am giving a relatively grammar free version of (7) that is simultaneously a description of the grammaticalized aspects of (7).
These redundant paraphrases are of the report type described by Harris 1969. In subsequent work Z. S. Harris described in what way could these redundant elements, within the framework of an applicative grammar, be classified according to the operators that selected them and how the analysis of the conditions of their deletion could lead to restore the grammar they came from.

In short, how the description of grammar can nearly coincide with an entirely linguistic process of deletion of the grammatical constraint. This phenomenon perfectly complies to Harry's statement that "every natural language must contain its own metalanguage, i.e. the set of sentences which talk about any part of the language, including the whole grammar of the language" (1968: 17) and that "Any adequate description of language has to provide for the metalanguage as a not immediately distinguishable part of the language" (Ibidem). It is also in accordance, with slight differences, with Wittgenstein's insights about the nature of human language (see Ibrahim 2001c). It stresses the fact that one of the main defining properties of human languages is their ability to be self-describing.

One can therefore say that if lexicon-grammar has achieved a method to describe and recognize grammatical constraints compatible with an explanation based on the existence in all languages of equivalence classes (Harris 1968: 147 footnote 35), the report method allows us to define these constraints via explanatory redundancies. At the same time, the applicative framework through which redundancies are reduced guarantees the pertinence of the mapped redundancies as well as the rebuilding of all deleted elements.

Nevertheless this way of defining grammatical constraints explains how an element of the lexicon is inserted in a minimal utterance but does not explain how and why this insertion is a source for the building of meaning. This way shows how the units of a sequence, an utterance or a chain of utterances are realized but it does not explain why every variation in the way of realizing these units brings up meaning.

An explanation becomes possible at this prospect only if, as we constructed an equivalence class grouping together the grammatical constraints and their defining redundancies, we construct an equivalence class grouping together the redundant definitions and the analytic configurations one can show they are analogous to the final semantic configurations the speaker recognizes.

The analytic configurations we call analytic defining matrices play towards the redundant definitions of grammar the same role the redundant definitions of grammar play towards elementary syntactic structures. These matrices have two functions: they altogether explain grammatical constraints and cancel them. Their redundancy is explanatory in the sense that it converts a grammatical constraint into words but in the same time it cancels the idiosyncratic constraint of the particular language it is describing. Actually, analytic matrices, although not universal as a material mapping of a given utterance, are universal regarding the principles governing their arrangement and the explanation they can give of this arrangement. In short, they can show by setting themselves as examples and without stating any rule but in a fairly precise way, how and why the type of deletable explanatory redundancies is not the same, for a given configuration, for instance, in English, French, Italian, Spanish, Standard Arabic, Egyptian Arabic or German.

Therefore, the analytic matrices gain necessarily a virtual extension, since for a given type of redundant matricial definition of explanation, all positions will not be necessarily filled in all languages, furthermore they can be filled differently if they account for two equivalent variants within the same language, although every position - be it filled or not - contributes to the construction of meaning.\footnote{A crystal clear phenomenon in children's acquisition of their mother tongue (See Ibrahim & Martinot 2004).}

For instance, if we try to build the defining matrix of the English sentences:

(8) Zizou kicked a goal
(8') Zizou scored a goal

in order, as well to understand what is going on in English as to fix a truly automatic translation device able to handle, for instance,

French:
(9) Zizou a marqué un but
(9') Zizou a rentré un but
(9") Zizou a mis un but

Italian:
(10) Zizou a segnato una rete
(10') Zizou a segnato un gol

Spanish:
(11) Zizou metió un gol
(11') Zizou marcó un gol

Standard Arabic:
(12) Sajalla Zizou hadafan\footnote{literally registered Zizou goal}.
Egyptian Arabic:
(13) Zizou gaab goon\footnote{literally Zizou brought goal}.

We should build a matrix of the type:\footnote{For a thorough discussion of the arguments for the construction of this matrix and the different types of reductions involved, through these examples plus German with a French starting point, see Ibrahim 2004b: 54-59.}

\begin{verbatim}
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
</table>
| A kick of / the foot / in / a given football
\end{verbatim}

\textbf{Source:} A kick of / the foot / in / a given football

We should build a matrix of the type:\footnote{For a thorough discussion of the arguments for the construction of this matrix and the different types of reductions involved, through these examples plus German with a French starting point, see Ibrahim 2004b: 54-59.}
The English expression (8) happens to include the nine positions corresponding to the nine components needed to actualize goal within this matrix. All the languages do not fill the same positions, neither does expression (8') in English, as can easily be seen when comparing example (8) with those of the other languages we have chosen. The matrix of an utterance in a particular language can be incomplete with regard to the universal virtual matrix without being incomplete by itself due to the very high redundancy of the universal matrix. Furthermore, the starting point of the construction of meaning can be quiet different event in the same language as we see in English where (8) starts from the first position while (8') starts from the last one.

The components of the universal matrix fall into a few functional categories that received a fairly good syntactic characterization and seem universal (Ibrahim 2004):
- nouns: The source of predication - In our examples it has to be goal or something that is part of it like italian rete.
- operators: as defined in all his work by Harris, that is elementary, indivisible and few elements that can freely produce meaning by applying to almost any other operator or argument according to their type. In our examples given - make.
- supports: as we defined them, following partly Maurice Gross and Anne Daladier, that is bearers of much of what a noun needs to act as a predicate but also of the type of appropriateness within aspectual, kinetic and perspective dependent domains needed by the actualizing process occurring when a noun enters in a discourse. The lexicon numbers more supports than operators but they are still a small number in all languages with regard to other functional categories. They can often be deleted - surface operators cannot -, always hold different transformational properties from their non-support homonyms, are indivisible, have no assignable semantic representation. Their high degree of appropriateness does not block their combination properties. In our examples kick, score.
- auxiliaries as they are traditionally defined.
- classifiers as we define them, that is labels for relatively homogeneous domains that play an important role in the building of the matrixes but are generally reduced in surface structures. In our examples ball.
- correlates as we define them, that is a small number of links between the nominal predicate and the discourse stating a very few recurrent elements needed for the description of all kinds of actualizing processes like state, shape, position, property, ... In our examples direction.
- generic substitutes as any one can define them. In our examples someone.
- specific substitutes
- quantifiers
- link markers (including copulas). We can also consider them as structure markers. In our examples in, to, by.

All the components of a matrix except the noun, converge and compete towards the actualization of the noun, considered as the home of the sentence predication whatever it is. This means that the Defining Matrix Analysis Model is a Redundancy Zeroing Device for the mapping and reduction fo the constraints described within the Lexicon-Grammar in a way that shows which of these constraints could be universally relevant and therefore eligible for a comprehensive survey through all the languages of the world so as to increase our chances for a genuine automatic data processing.

References:
Ibrahim, Amr H.

This is of course, regarding English, a mere coincidence that has nothing to do with what could be seen as a property of this language although we did not choose this example by chance. We considered the fact that football was born in Great Britain and grown in an English speaking community.
2005a, "Le paradigme des supports de point de vue en français et en arabe", Les périphrases verbales, (Hava Bathe-Shyldkrot & Nicole Le Querler éds), Amsterdam / New York: John Benjamains, Col. LIS.