clacCÍRCULO
de lingüística aplicada a la comunicación

65/2016

TOWARDS THE COMPUTATIONAL IMPLEMENTATION OF ROLE AND REFERENCE GRAMMAR: RULES FOR THE SYNTACTIC PARSING OF RRG PHRASAL CONSTITUENTS

Francisco José Cortés-Rodríguez

Instituto Universitario de Lingüística 'Andrés Bello', Universidad de La Laguna

fcortes *en* ull edu es

Abstract

The goal of this work is twofold: on the one hand, it seeks to update the description of phrasal constituents in Role and Reference Grammar; on the other, it aims at providing a computational implementation of such structures within the Grammar Development Environment, a component of ARTEMIS ("Automatically Representing Text Meaning via an Interlingua-Based System"), a Natural Language Processing prototype developed with the aim of binding natural language fragments with their corresponding grammatical and semantic structures.

In order to attain both tasks the analysis focuses specifically on the design of the rules that would account for the linkage between the syntactic and the semantic

Francisco José Cortés-Rodríguez. 2016.
Towards the computational implementation of role and reference grammar: rules for the syntactic parsing of RRG phrasal constituents
Círculo de Lingüística Aplicada a la Comunicación 65, 75-108.
http://www.ucm.es/info/circulo/no65/cortes.pdf
http://revistas.ucm.es/index.php/CLAC
http://dx.doi.org/10.5209/rev_CLAC.2016.v65.51982

representations of Referential Phrases, as proposed initially in RRG. This proposal involves a reinterpretation of the constituents and operators in the Layered Structure of the RP taking into account that in ARTEMIS the assignment of a syntactic-semantic structure to a given sequence is based on the activation of grammatical rules plus a set of Attribute Value Matrixes related to the grammatical features of the constituents of RPs.

Keywords: Role and Reference Grammar, Referential phrases, ARTEMIS, parsing rules, attribute value matrixes

Contents

1. Introduction

A growing concern for many grammatical models, especially functionally oriented theories, is the development of the conditions to satisfy what may be called the criterion of computational adequacy.[1] In line with this interest, some recent contributions (Guest 2009; Salem, Hensman and Nolan 2008; Diedrichsen 2014) within Role and Reference Grammar (henceforth RRG; Van Valin and LaPolla 1997; Van Valin 2005; Pavey 2010)

[1] This work has been developed within the framework of the research project “Desarrollo de plantillas léxicas y de construcciones gramaticales en inglés y español. Aplicación en los sistemas de recuperación de la información en entornos multilingües” (FF12011-29798-C02-02), funded by the Spanish Ministry of Science.

have been devoted to the development of this issue. However, it is still a challenge to devise a fully-fledged syntax-semantics interface which is computationally implemented.

Works like Van Valin and Mairal Usón (2013) or Periñán-Pascual (2013, 2014) draw the guidelines of a system of natural language understanding –ARTEMIS- based on the theoretical tenets of this grammar. ARTEMIS ("Automatically Representing Text Meaning via an Interlingua-Based System") has been implemented as a parsing device within FungramKB ('Functional Grammar Knowledge base") for the computational treatment of the syntax and semantics of sentences.

The goal of this work is twofold: on the one hand, it seeks to update the description of phrasal constituents in RRG, taking into account Van Valin's (2008) programmatic proposal, which brings about some significant variations in the interpretation of phrasal constituents as they were originally described in Van Valin and LaPolla (1997) and Van Valin (2005); on the other, it aims at providing a computational implementation of such structures within the Grammar Development Environment (GDE), a component of ARTEMIS whose task is to provide an effective parsing of morphosyntactic structures. Despite the fact that this NLP prototype is based on the theoretical tenets of RRG (Van Valin and Mairal Usón 2013; Periñán-Pascual 2014, Periñán-Pascual and Arcas-Túnez 2014), a reinterpretation of some of the constituents and operators in the Layered Structure of Referential Phrases (henceforth RPs) is necessary in order to comply with the requirements of the syntactic parser, taking into account that in ARTEMIS the assignment of a syntactic-semantic structure to a given sequence is based on the activation of grammatical rules plus a set of Attribute Value Matrixes (AVMs) related to the grammatical features of the constituents of RPs.

The remainder of the paper is organized as follows: Section 2 offers a brief description of the interrelation among RRG, FunGramKB and ARTEMIS highlighting both the points of consensus and the adjustments required for an effective integration of RRG functional syntax into the Grammar Development Environment (GDE) which will perform the parsing operations within ARTEMIS. Section 3 updates the syntactic analysis of phrasal constituents in RRG by fully developing the programmatic proposal put forward in Van Valin (2008), which starts from a radical substitution of the analysis based on formal categories such as NP or AdjP for a functionally oriented description in terms of the categories Referential Phrase (RP) and Modifier Phrase (MP). In section 4 a description of the elements necessary for the computational parsing of these new categories within the GDE is proposed. After a conclusion section and the list of references, the paper is rounded off with an appendix including a list of the abbreviations that have been used in the parsing rules proposed and in the AVMs.

2. Artemis, RRG and FunGramKB

As stated above, ARTEMIS is an NLP prototype designed primarily for natural language understanding. One crucial differentiating feature with regard to other NLP systems with similar tasks is the fact that ARTEMIS seeks to be linguistically motivated. This involves adopting a linguistic theory, RRG, as a foundational pillar for the construction of the components of the NLP system. Periñán-Pascual (2013: 209) explains some features of RRG that make it suitable for NLP:

(a) RRG is a model where morphosyntactic structures and grammatical rules are explained in relation to their semantic and communicative functions.

(b) RRG is a monostratal theory, where the syntactic and semantic components are directly connected through a bidirectional “linking algorithm”.

(c) RRG is a model that makes strong claims to typological adequacy.

However, given the conditions imposed by NLP environments, a direct computational “translation” of RRG’s grammatical structures and rules is impossible, and some fine-tuning becomes necessary. These adjustments involve, on the one hand, resorting to other models that complement RRG in an area where this grammar is still underdeveloped. I am referring to semantic representations; in fact, the contributions within the Lexical Constructional Model (LCM; Ruíz de Mendoza and Mairal Usón 2008; Mairal Usón and Ruíz de Mendoza 2008, 2009) have become significantly relevant to develop a fully-fledged system of semantic representations.

On the other hand, ARTEMIS deploys FunGramKB knowledge base (Periñán-Pascual and Arcas-Túnez 2004, 2005, 2007, 2008, Periñán-Pascual and Mairal Usón 2009, 2010,2011), which in turn also exploits constructional schemata as a system for meaning representation.

FunGramKB includes the following interrelated modules:

(a) The lexical component, which is language-specific and consists of two submodules: the lexicon (which includes in the format of entries all the linguistic information related to the lexical units) and the morphicon (which deals with all inflectional processes of a language).

(b) The grammatical level, also language dependent where constructional schemata of a given language are stored. Such schemata are organized according to the 4 layer classification proposed within the LCM.

(c) The Conceptual component is language independent and stores all deep semantic units and structures into different submodules: the ontology (a hierarchical storehouse for concepts in a human mind), the cognicon (or repository of procedural conceptual schemas or scripts to encode stereotypical actions) and the onomasticon (for real world entities and events).

One pivotal element that binds FunGramKB with RRG and the LCM is the integration of the LCM’s constructional templates and RRG’s lexical representations (logical structures) into the knowledge base’s language-independent formalism for text meaning representation, namely Conceptual Logical Structures (CLS). CLS takes as a backbone for semantic representation the Aktionsart characterization of lexical units as encoded in the Logical Structures of RGG. To show the differences between both types of constructs, Periñán-Pascual (2013: 218) offers the semantic representation of the sentence “The juice froze in the refrigerator” first in its RRG version:

(1) $<_{IF}$ DEC $<_{TNS}$ PAST 〈 **be-in’** (refrigerator, [[**do’** (juice, [freeze’ (juice)])] CAUSE [BECOME **black’** (juice)]) >>

and next, the corresponding FunGramKB CLS:

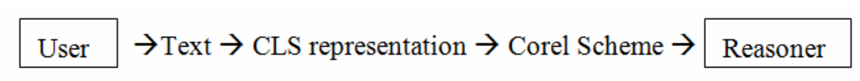
(2) $<_{IF}$ DECL <Tense past <CONSTR-L1 RESI <CONSTR-L1 INCH AKT ACC +FREEZE_00 (+JUICE_00-Referent, +BLACK_00Result)] (+REFRIGERATOR_00-Location) 〉 >>>>

There are significant changes between (1) and (2) (Periñán-Pascual 2013: 218-219), among which the most relevant is the substitution of predicates by concepts from FunGramKB’s ontology (as are +JUICE_00 and +BLACK_00 in the above example).

Although the CLS brings a heavier conceptual load into semantic representations, it needs still some refining from a computational perspective. In fact, for reasoning purposes it is necessary to model CLS representations into COREL structures[2]. Thus the CLS in (2) is modeled into a COREL scheme of the following type:

(3) +(e1: past +FREEZE_00 (x1)Theme (x2: +JUICE_00)Referent
(f1: (e2: +BECOME_00 (x2)Theme (x3: +BLACK_00)Attribute))Result
(f2: +REFRIGERATOR_00)Location)

In simple terms, the process involved in understanding a stretch of natural language with the tools that have been described is summarized in the following figure:



The main goal of ARTEMIS is to deal with all the phases of this process; consequently, it consists of the following modules: The Grammar Development Environment (GDE), the CLS Constructor and the COREL-Scheme Builder. As can be inferred from the previous description of examples (1-3), the two last modules are designed to transfer the shallow semantic representations of sentences into conceptually deeper structures amenable for their processing in different PLN tasks. However, prior to this it is necessary to make an effective computational parsing of the morphosyntactic structure underlying sentences based on the principles of RRG for grammatical descriptions; this is the goal of the GDE.

The analysis carried out in this paper seeks to enrich the information that is required for an effective parsing of the so-called RPs in RRG, and which would become a part of the GDE. So far, the GDE consists of two basic types of theoretical constructs, a set of rules that account for syntactic structures and a library of Attribute-Value Matrixes (AVMs) for grammatical units.

[2] COREL (COnceptual REpresentation Language) is an interface language to formalize conceptual knowledge into FunGramKB. For a description of the grammar of this notational language see Periñán Pascual and Mairal Usón (2010).

Rules can be divided into the following types: (a) syntactic rules whose task to account for the generation/recognition of the underlying Layered Structure of a Clause (LSC); (b) constructional rules which guide the embedding of the structure of argument-predicate constructions (L1 Constructions in the LCM, such as resultatives, caused motion constructions, etc.) into the enhanced LSC; and (c) lexical rules to tokenize the abstract features encoded in the LSC by resorting to the information stored in the Lexicon and the Ontology of FunGramKB.

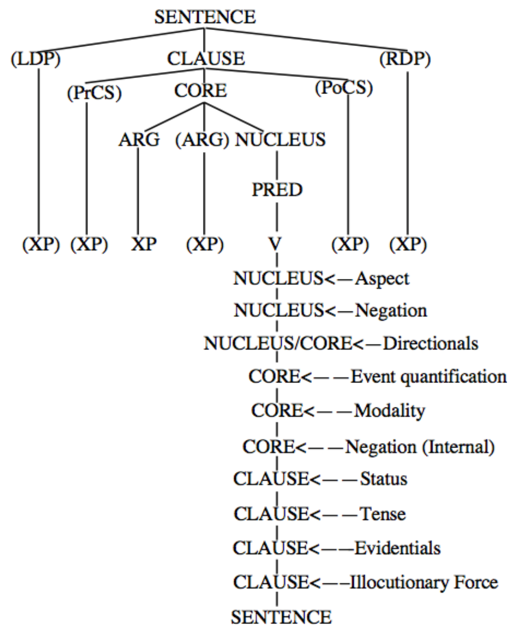
It is important to highlight the fact that grammatical constituents in the LSC are not atomic, but they are feature-bearing structures, namely AVMs. Thus, ARTEMIS also pays a tribute to unification approaches (Boas and Sag 2012; Sag, Wasow and Bender 2003) and parsing operations are guided not only by phrase-structures rules but also by constraint-checking operations on the selectional and semantic information encoded in the AVMs. An example of an AVM for the syntactic node CORE which forms part of the Layered Structure of Clauses would be the following[3]:

(4)
```
<Category Type="CORE">
<Attribute ID="Num" />
<Attribute ID="Template" />
<Attribute ID="Tense" />
<Attribute ID="Neg” />
<Attribute ID="Mod" />
</Category>
```

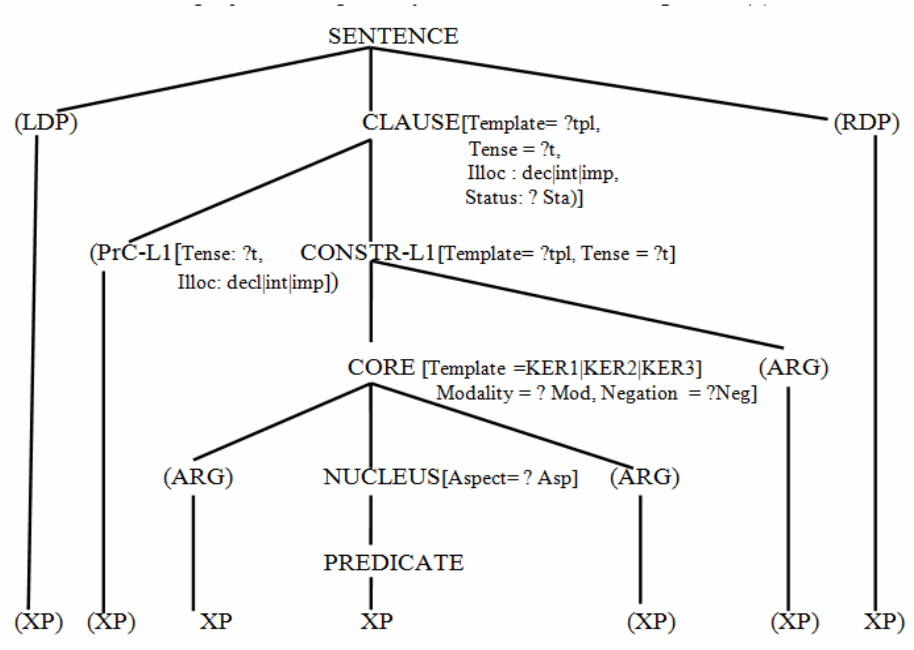
The inclusion of unification devices in the prototype also leads to a fundamental adjustment of the original RRG syntactic analysis: the so-called operator projection (operators are the grammatical categories such as tense, mood, aspect, etc. which modify the layers of the clause) is overridden mainly by the AVMs, where the grammatical features that modify the different categories are lodged. Thus, the original syntactic pattern (including constituent and operator projections) for simple clauses in RRG:

[3] AVMs are encoded in XML format, similar to that of other platforms for the analysis of human language data, as is NLTK (Natural Language Toolkit; Bird, Loper and Klein 2009). See the next section, for a detailed explanation of the information encoded in this AVM.

(5)



will look differently once the operator projection is substituted by feature-bearing nodes in the constituent projection, as partially reflected in the tree diagram in (6)[4]:



[4] The diagram is obviously incomplete, since neither all the nodes are endowed with their corresponding AVMs nor the AVMs which are represented include all its possible attributes; they just represent some of the 'prototype syntactic rules' for clauses proposed as preliminary within ARTEMIS (the so-called 'version 1.0' rules). Further research on sentences and other syntactic structures is necessary to fully develop the set of necessary rules within the GDE.

3. The scope of analysis: description of grammatical processes and its treatment in RRG

Van Valin (2008) introduces some important variations in the interpretation of RRG syntactic categories as originally described in both Van Valin and LaPolla (1997) and Van Valin (2005). Such changes are based on the assumption that the most significant syntactic categories are not projections of a lexical head, but projections of the functional status of constituents within the clause. This is already captured in the grammatical components of clauses in RRG, as evidenced by the existence of categories such as Nucleus (NUC), CORE and Periphery (PER). However, at the level of phrasal categories there was still a certain inconsistency, as labels such as NP or AdjP are not functionally motivated. Thence two types of constituents are proposed to account for a functional and typologically valid treatment of such lexical projections[5]: Referential Phrases (RPs) and Modifier Phrases (MPs). From this it follows that the description provided in Van Valin and LaPolla (1997: 52-67) and Van Valin (2005: 21-30) for NPs must be adapted now to RPs, and that the layered structure assigned to NPs is now attributable to RPs with the necessary adaptations. The remainder of this section will then offer a description of what we may call the Layered Structure of Referential Phrases (LSRP), and which will replace Van Valin and LaPolla's (1997) and Van Valin's (2005) Layered Structure of Noun Phrases.

The fact that RPs are described in terms of a layered structure in RRG is due to their remarkable structural equivalence with clauses. The most striking similarity is that nouns (and adjectives) can take arguments, as verbs in clauses do. This is especially relevant in the case of derived nominals as in *the destruction of the city by the earthquake* and of the relational nouns like *friend* or *relative*, commonly followed by PPs introduced by *of* which can be considered as their direct arguments, following Nunes (1993). It is important to highlight that, in accordance with Van Valin (2008: 167) the parallelism between RPs and clauses increases because the former need not be endocentric categories; i.e. there is no restriction for RPs to be headed exclusively by any specific lexical category. The same as there is a strong tendency, but not an absolute correlation, for verbs to be the nuclei of clauses, there is a strong tendency for nouns to be the nuclei of RPs, but it is not always the

[5] With regard to the typological aspects, Van Valin (2008) also proves the fact that there are no universally valid typological categories beyond the fundamental categories of noun and verb. Furthermore, not all languages align consistently morpholexical categories with functional distinctions. For example, in Nootka (cf. Van Valin 163-164) it is feasible to align nouns with a predicational function and verbs with a referring function, as happens in the Nootka sentence in example (7).

case that there is a nominal nucleus. Van Valin (2008: 163-164) offers examples from different languages, as are Tagalog and Nootka:

(7) Qo:□as-ma wa□a:k-□i (Nootka)

Man-3SGPRES go-the ('the one going is a man')

In this sentence, a verb is used as an argument and a noun is used as a predicate. Similar cases are found in languages which are more familiar to us, as is the case of German (Van Valin 2008: 167)

(8) Der Lange ist eingeschalafen

The.M.Sg.NOM tall be.3SGPRES fall asleep.PASTPART ('The tall one has fallen asleep')

Or even Spanish:

(9) Los ricos también lloran

The.M.Pl. rich.M.Pl. also cry ('The rich also cry')

A preliminary general scheme of the LSRP is given in the following figure[6]:

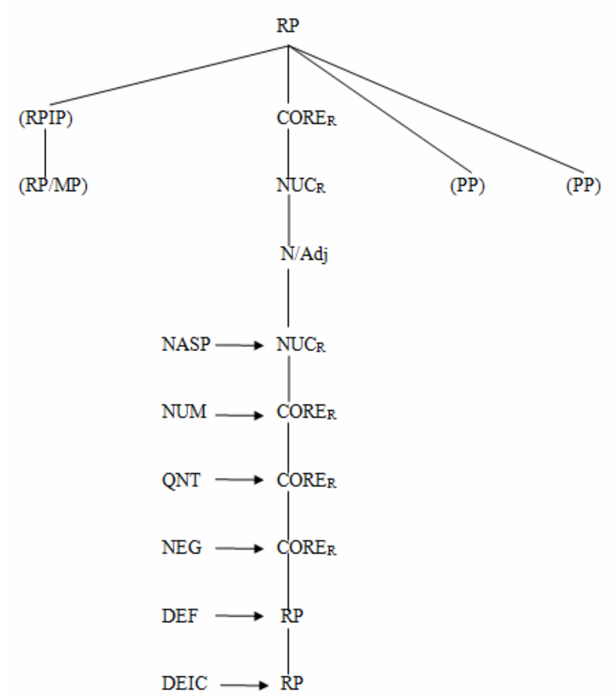
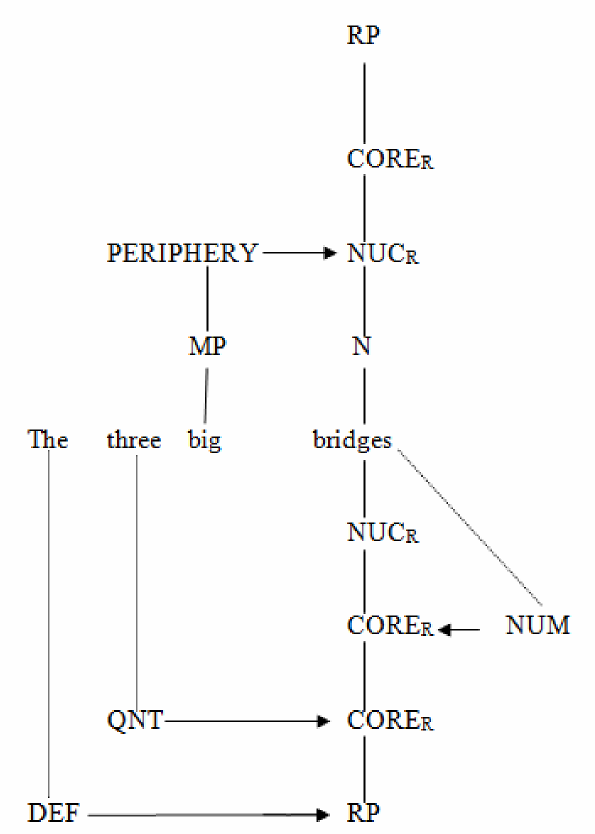


Figure 1: The Layered Structure of Relational Phrases

[6] NASP = Nominal aspect; NUM = Number; QNT = Quantification; NEG = Negation; DEF = Definiteness; DEIC = Deixis.

The following examples have been borrowed from Van Valin (2005: 25) and adapted to the new format of analysis in RRG:

(10) The three big bridges



(11) The construction of the bridge by the company in New York City

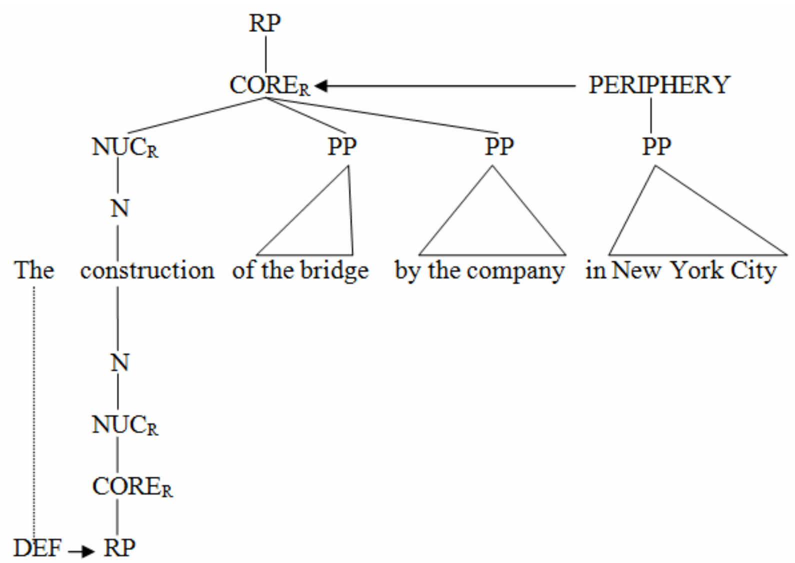


Figure 1 shows that each of the layers in RPs can be modified by different operators. The whole set of operators in the LSRP is given in the following table (adapted from Van Valin 2005: 24):

Nuclear Operators
    Nominal aspect (count-mass distinction, classifiers in classifier language)
Core Operators
    Number
    Quantification (quantifiers)
    Negation
RP Operators
    Definiteness
    Deixis

Table 1- Operators in the LSRP

As in the case of clauses, the ordering of operators is constrained and responds iconically to their respective scopes (cf. Rijkhoff 1990, 1992); there is also a periphery for each layer, as happens in clauses as well. The nuclear periphery would host restrictive MPs and restrictive relative clauses[7]:

(12) My *dear old wood* hammer *that never lets me down*

The core periphery would include setting PPs and MPs as *in New York City* in example (11) above, and non-restrictive modifiers (relative clauses and appositional phrases) will occupy the RPs periphery as is the RP *a cupcake expert* in *Rebeca*, *a cupcake expert*. Example (11) also reveals that PPs can occupy two different positions within the LSRP, since they can be function within the CORE as a kind of exocentric RP, behaving as an argument of the NUC; or they can be treated as peripheral elements, thus having a predicative setting function.

[7] AdjPs are probably the category that has undergone the most significant re-analysis in RRG; initially they were considered nuclear operators, then in Van Valin (2005) were integrated in the constituent projection of NPs as peripheral modifiers in order to provide them with a status similar to the rest of lexical categories, and finally Van Valin (2008) aligned them as MPs together with adverbial phrases.

Nunes (1993) shows that NPs have one direct core argument, which is marked by the preposition *of*. Therefore, *of* is non-predicative in these structures and it is semantically empty; this is proved by the fact that it occurs with a whole range of semantic functions, as in the following examples (Van Valin and LaPolla 1997: 55):

(12)	The attack of the killer bees	Agent
(13)	The gift of a new car	Theme
(14)	The destruction of the city	Patient
(15)	The leg of the table	Possessor
(16)	The resupplying of the troops	Recipient

On the other hand, predicative prepositions have semantic content, as the rest of predicates. Examples (27) and (28) in Section 4 show the different structural configuration of predicative and non-predicative PPs.

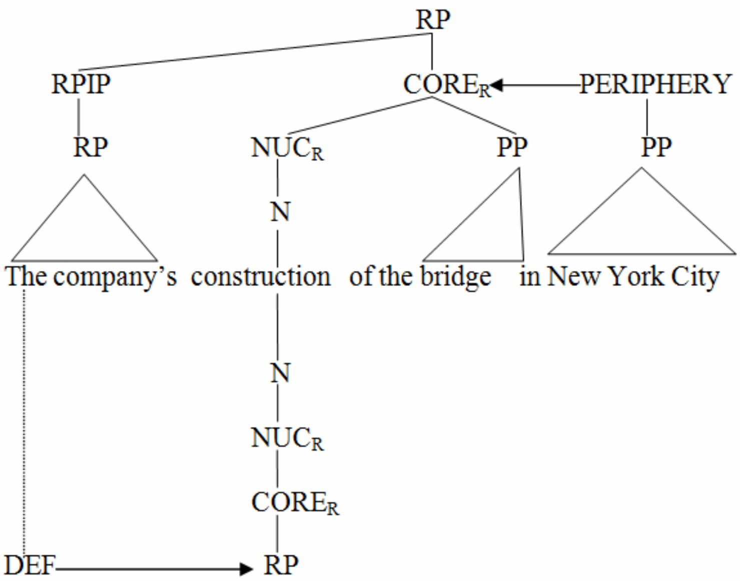
With regard to Modifier Phrases (MPs), they also have an internal layered structure. Van Valin (2008: 172) gives two reasons to support this: firstly, there are languages with MPs that have as a nucleus an Adjective that takes a core argument. Van Valin (2008: 172) mentions German as one of these languages (e.g. *der auf seinen Sohn stolze Vater* 'the of his son proud father'), and states that English does not allow such type of structures (**the proud of his son father*); however, this impossibility occurs only in the case of attributive adjectival MPs, but it is possible to have MPs with a core argument if they appear within an RP in a post nuclear position; for example: *a speech redolent of old-aged morality*. Secondly, the modifiers in MPs can themselves be modified; this involves that the MP must make room for a periphery to house the modifying MP; e.g. *a very brightly debated proposal*.

There is an especially relevant slot in RPs, the so-called 'RP Initial Position' (RPIP) that exhibits features similar to those of both the detached (f.i. it allows Genitive MPs as in *Today's class*) and the precore slot positions of clauses (wh-words can also occupy this position, as in *which evening dress*)[8]. The RPIP plays a role both in the constituent and

[8] There is also an RP Final position, which is relevant for other languages; for example in the RP *Wówapi ki lé* 'book the this' from Lakhota the demonstrative is in RPFP. Something similar occurs to the article in Mparntwe Arrenrnte, as in *kngwelye nhenhe re* 'dog this 3SGDEF'.

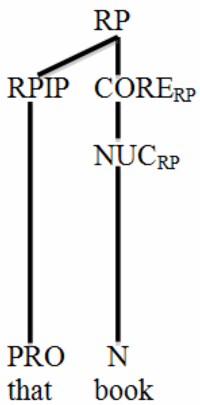
the operator projections in English as it is the place where the definiteness operators are located:

(17) The company's construction of the bridge by the company in New York City (adapted from Van Valin 2005: 27)

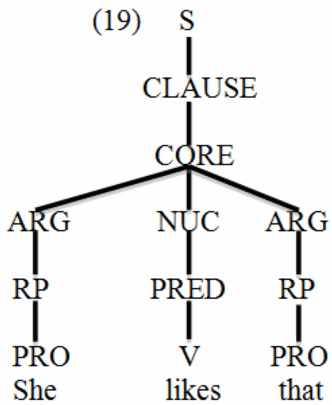


English possessive and demonstrative modifiers also occupy this position; however, their equivalent pronouns would function as RPs on their own and they do not have a layered structure, as happens with the rest of pronouns and proper nouns (cf. Van Valin and LaPolla 1997: 56). Compare the following (adapted from Van Valin and LaPolla 1997: 62):

(18)



(19)



4. Rules and AVMs implementation within ARTEMIS

In its present state of development, still as a proof-of-concept NLP system, ARTEMIS needs a fully-fledged development of many specific syntactic rules and AVMs to satisfy the system for grammatical analysis based on RRG. This is especially true in the case of phrasal structures: whereas the rules for the analysis of constituents at clause level is already partially based on the layered structure of the clause, as evidenced by the existence of nodes such as CORE, PER o NUC,[9] the internal structure of phrasal constituents does not contemplate the way they are analyzed in RRG. In this section I will provide a first attempt at developing the syntactic rules and the AVMs necessary for this NLP prototype to analyze phrasal constituents in compliance with RRG's conception.

The first changes to be introduced will affect the overall structure of the clause, as I intend to further integrate the RRG syntactic perspective into ARTEMIS. Thus, the initial rules in version 1.0[10]:

> S -> CL
> CL[Tense=?t, Template=?tpl] -> CONSTR-L1[Tense=?t, Template=?tpl] || CONSTR-L1[Tense=?t, Template=?tpl] PER

will be replaced by the following ones

> S —> CL|| LDP CL|| CL RDP

This new rule introduces the node tags LDP and RDP, to account for the positions of detached constituents, which are clause external but sentence internal elements.

> CL [Template= ?tpl, Tense = ?t, **Illoc : dec**|**int**|**imp, Status: ? Sta**)] -> CONSTR-L1[Template= ?tpl, Tense = ?t] || CONSTR-L1[Template= ?tpl, Tense = ?t] PER || **AUX** [Tense: ?t, Illoc: int|imp] CONSTR-L1[Template= ?tpl, Tense = ?t] || AUX [Tense: ?t, Illoc: int|imp]-L1[Template= ?tpl, Tense = ?t] PER ||

[9] However, the same does not hold either for the structures above the clause and below the sentence (note that there isn`t any rule to account for precore positions or detached constituents).

[10] As mentioned in note 3, version 1.0 rules in ARTEMIS are only a kind of 'toy grammar' and, therefore, they are still susceptible to significant modifications, as further extensive research is required to develop it as a fully-fledged parser.

PreC-L1 AUX [Tense: ?t, Illoc: decl|int|imp] CONSTR-L1[Template= ?tpl, Tense = ?t] || **PreC-L1** AUX [Tense: ? t, Illoc: int|imp] CONSTR-L1[Template= ?tpl, Tense = ?t] PER

The rule that expands the CL node also contemplates the introduction of two nodes, namely AUX (for auxiliary verbs) and PreC-L1. They are necessary to account for those cases in which there is a clause initial auxiliary verb that in turn can be preceded by a constituent in a clause initial position. In the original RRG analysis this is described as a PreCore slot, and it is the place typically occupied by question words in languages in which they do not appear in situ, and also by fronted constituents as in *To this list a thousand more complaints could be added*. However, since the node CONSTR-L1 has been introduced to account for the embedding of predicate-argument constructions from the (L1-Constructicon), it seems more adequate to re-analyze the Precore as a PREC-L1; this will allow us to account for the constituents that appear in that position and which are introduced in the LCS by means of constructional rules, as in the following examples:

(20) *For whom* did you buy the cake? (Beneficiary RP in PreC-L1 subsumed in the CLS from a Beneficiary L1 Construction)

(21) *Where* did the cork float into? (Interrogative PRO constituent introduced by the L1 CMOT Construction into the CLS)

AUX in initial position helps to explain the structure of yes/no questions and imperative clauses (*Do come here! Don't talk*!)[11]. This is the reason why the value 'decl' for the Illocutionary force attribute is not available in the AVM of the node AUX.

Another interesting variation affects the Attributes associated to the different nodes involved in the previous rules. This must be done in order to capture the whole range of operators that affect the different layers in syntactic representations. For example, the CL node will now have the following AVM:

[11] For a detailed description of the role of AUX items in the GDE, cf. Díaz Galán and Fumero Pérez (2015), which also offers a first description of some rules and AVMs for the syntactic parsing within ARTEMIS of clause-level grammatical operations involving do-auxiliary insertion.

```
<Category Type="CL">
        <Attribute ID="Illoc" />
        <Attribute ID="Status" />
        <Attribute ID="Template" />
        <Attribute ID="Tense" />
</Category>
```

And the attributes “Status” and “Illocutionary Force” would have the following AVMs:

```
<Attribute ID="Status " obl="*" num="1">
        <Value>?sta</Value>
        <Value>possibility</Value>
        <Value>certainty </Value>
        <Value>inference </Value>
</Attribute>

<Attribute ID="Illoc " obl="*" num="1">
        <Value>?illoc</Value>
        <Value Tag=”declarative” >dec< /Value>
        <Value Tag=”interrogative” >int< /Value>
        <Value Tag=”imperative” >imp< /Value>
</Attribute>
```

The following significant changes to be introduced in ARTEMIS repository of syntactic rules would be the substitution of the node tags for the categorial phrases NP, AdvP and AdjP AVMS for the corresponding RP and MP tags; this would account for the ‘external syntax’ of RPs and MPs. These changes affect the so-called “argumental”, “nuclear” and “peripheral” levels (levels 4, 3 and 2 in version 1.0). Strictly speaking, the nodes ARG and ADJUNCT are no longer used in the latest version of RRG, but the analysis of their usefulness in ARTEMIS lies beyond the scope of this paper. Suffice by now to consider that their preservation helps to keep separate the AVMs for these constituents and their corresponding attributes as members of the clause (i.e. the kind of thematic role they fulfill) from those pertaining to their internal configuration at phrase level (as is for example the definiteness of RPs). In keeping with this line of reasoning, the attribute of number has been erased from the AVMs of ARGs and ADJUNCTs.

Another issue to be considered in further analyses is the (in)adequacy of including an Argument-Adjunct (AAJ) node tag, a type of constituent proposed almost exclusively by RRG. This is the case of Instrument constituents form the Instrument-as-Subject Construction which are embedded in the CLS by the constructional rules in ARTEMIS, and it would affect –among others- the rule:

ARG[Var=w, Phrase=RP, Num=?n, Role=Instrument, Template=INST]-> RP[…]

Again, this issue also is beyond the scope of our study.

These changes are reflected in the following set of rules existing in the prototype proposal in ARTEMIS:

#Level 4: Argument Level
ARG[Var=x, Role=Agent|Theme, Template=KER1|KER2] -> **RP**
ARG[Var=y, Role=Theme|Referent|Location|Goal, Template=INCH|KER2|MIDD|RESI] -> **RP**
ARG[Var=y, Phrase=RP, Role=Goal|Instrument, Template=INST] -> **RP**
ARG[Var=y, Phrase=PP, Prep=?p, Role=Theme|Referent|Location|Goal, Template=KER2] -> PP[Prep=?p]
ARG[Var=w, Role=Instrument, Template=INST] -> **RP**
ARG[Var=w, , Role=Manner, Template=MIDD] -> **MP**
ARG[Var=w, , Role=Goal, Template=CMOT] -> PP[Prep=?p]
ARG[Var=w, Role=Instrument, Template=INST] -> PP[Prep=?p]

#Level 3: Nuclear Level
NUC[Tense=?t, Num=?n, Template=?tpl] -> PRED[Tense=?t, Num=?n, Template=?tpl]
PRED[Tense=?t, Num=?n, Template=?tpl] -> VERB[Tense=?t, Num=?n, Template=?tpl]
NUC-S[Role=Result, Template=RESI|RESU] -> PRED-S[Phrase=**MP**]
NUC-S[Prep=?p, Role=Result, Template=RESI|RESU] -> PRED-S[Phrase=PP, Prep=?p]
PRED-S-> **MP**
PRED-S[Prep=?p] -> PP[Prep=?p]

#Level 2: Periphery

PER -> ADJUNCT || ADJUNCT ADJUNCT || ADJUNCT ADJUNCT ADJUNCT

ADJUNCT-> **MP**

ADJUNCT[Prep=?p] -> NUC-A[Prep=?p]

NUC-A[Prep=?p] -> PRED-A[Prep=?p]

PRED-A[Prep=?p] -> PP[Prep=?p]

There are other changes affecting the attributes associated to the RP nodes in the rules above, since their AVMs must be significantly enriched with the attributes and values corresponding to the operators encoded in the LSRP, and described in the previous section. This is done below.

In order to capture the 'internal syntax' of RPs it is necessary to formulate the rules belonging to the last level in the syntactic rules of ARTEMIS, the Level 1 for phrasal realizations. They incorporate the rules necessary to adequately predict the layered internal configuration of RPs and MPs as explained in Section 3. Thus, the rules proposed for this level would be:

#Level 1: Phrasal structures

[1.a] Prepositional Phrases

PP-> PREP RP|| PER-PP CORE-PP

PER-PP -> MP

CORE-PP -> NUC-P RP

NUC-P-> PRED

PRED -> PREP

[1.b] Referential Phrases

RP -> RPIP CORE-RP PER-RP || RPIP CORE-RP || CORE-RP PER-RP || CORE-RP || PRO || PROD || PROP || PROQ || NOUX

RPIP -> PART ART || PART DETP || PART DETD || ART || DETP || DETD || RP|| MP

CORE-RP-> NUC-RP ARG-RP ARG-RP ARG-RP PER-CRP || NUC-RP ARG-RP ARG-RP ARG-RP || NUC-RP ARG-RP ARG-RP PER-CRP || NUC-RP ARG-RP ARG-RP ||
NUC-RP ARG-RP || NUC-RP || NUC-RP ARG-RP PER-CRP || NUC-RP PER-CRP

ARG-RP-> PP || CL

NUC-RP-> N || ADJ || ADJ PER-NRP || PROD || PROP || PROQ || NUMC || NUMO || DETQ PER-NRP N PER-NRP|| DETQ PER-NRP N || DETQ N PER-NRP || DETQ N || PER-NRP N PER-NRP|| PER-NRP N || N PER-NRP || NUMC PER-NRP N PER-NRP|| NUMC PER-NRP N || NUMC N PER-NRP || NUMC N || NUMO PER-NRP N PER-NRP|| NUMO PER-NRP N || NUMO N PER-NRP || NUMO N || NUMO NUMC PER-NRP N PER-NRP|| NUMO NUMC PER-NRP N || NUMO NUMC N PER-NRP || NUMO NUMC N || NUMC NUMO PER-NRP N PER-NRP|| NUMC NUMO PER-NRP N || NUMC NUMO N PER-NRP || NUMC NUMO N

PER-RP->MP || CL
PER-CRP->PP || MP || PP PP || PP MP || MP PP
PER-NRP-> MP || MP MP || MP MP MP || CL[12]

[1.c] Modifier Phrases

MP -> PER-MP CORE-MP || CORE-MP
CORE-MP-> NUC-MP || NUC-MP ARG-MP
NUC-MP - > ADJ || ADV || N || RP || CL || S
ARG-MP -> PP || CL
PER-MP -> MP

[12] Even though this rule allows only for 3 MPs in the Nuclear periphery of RPs, it is actually possible to have an indeterminate number of these phrases; however it is very uncommon to have as many as 4 or more.

The first five rules in the set above account for the internal configurations of PPs:

[1.a] Prepositional Phrases

PP-> PREP RP|| PER-PP CORE-PP

PER-PP -> MP

CORE-PP -> NUC-P RP

NUC-P-> PRED

PRED -> PREP

Following the structural distinction in RRG between predicative and non-predicative PPs, it is necessary to devise firstly a disjunctive rule like this:

PP-> PREP RP|| PER-PP CORE-PP

The first option accounts for non-predicative PPs whose preposition is licensed by the verb in the clause or the lexical head of phrases where such a PP functions as argument, and not by the preposition, as happens in the following examples:

(22) I showed the pictures to my neighbor

(23) The ambassador presented the new queen with a huge diamond

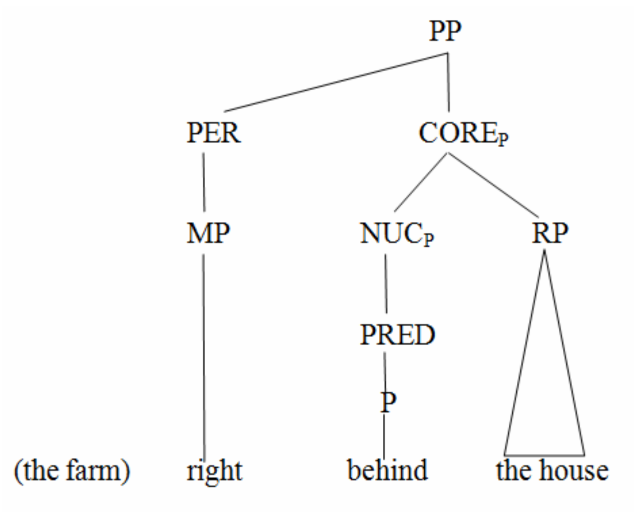
(24) The discovery of Mars

However, in a sentence like:

(25) I showed the pictures to my neighbor in the museum

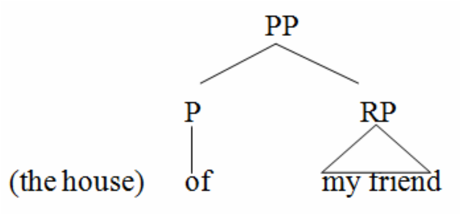
The PP in the museum has a predicative nucleus, the preposition in, which licenses its object. In these cases, the PP has a layered structure. For example, the PP right behind the house is analyzed in the following way:

(27)



Compare it with the non-predicative PP to my friend:

(28)



The first subset of rules above captures both types of PPs, especially the complex layering of predicative PPs.

The second subset of rules in the layer of phrasal constituents is devoted to spelling out the internal configuration of RPs in English:

(1.b) Referential Phrases

RP -> RPIP CORE-RP PER-RP || RPIP CORE-RP || CORE-RP PER-RP || CORE-RP || PRO || PROD || PROP || PROQ || NOUX

RPIP -> PART ART || PART DETP || PART DETD || ART || DETP || DETD || RP || MP

CORE-RP-> NUC-RP ARG-RP ARG-RP ARG-RP PER-CRP || NUC-RP ARG-RP ARG-RP ARG-RP || NUC-RP ARG-RP ARG-RP PER-CRP || NUC-RP ARG-RP ARG-RP || NUC-RP ARG-RP || NUC-RP || NUC-RP ARG-RP PER-CRP || NUC-RP PER-CRP

ARG-RP-> PP || CL

NUC-RP-> N || ADJ || ADJ PER-NRP || PROD || PROP || PROQ || NUMC || NUMO || DETQ PER-NRP N PER-NRP|| DETQ PER-NRP N || DETQ N PER-NRP || DETQ N || PER-NRP N PER-NRP|| PER-NRP N || N PER-NRP || NUMC PER-NRP N PER-NRP|| NUMC PER-NRP N || NUMC N PER-NRP || NUMC N || NUMO PER-NRP N PER-NRP|| NUMO PER-NRP N || NUMO N PER-NRP || NUMO N || NUMO NUMC PER-NRP N PER-NRP|| NUMO NUMC PER-NRP N || NUMO NUMC N PER-NRP || NUMO NUMC N || NUMC NUMO PER-NRP N PER-NRP|| NUMC NUMO PER-NRP N || NUMC NUMO N PER-NRP || NUMC NUMO N

PER-RP->MP || CL

PER-CRP->PP || MP || PP PP || PP MP || MP PP

PER-NRP-> MP || MP MP || MP MP MP || CL

The first of these rules establishes a distinction between two types of RPs: those that lack an internal layering, as happens with Pronouns when they appear alone and with Proper nouns, and the more complex ones, those in which it is necessary to distinguish three types of daughter nodes: the CORE-RP, the RPIP and the PER-RP. As it expressed in the rule, only the CORE-RP is not optional.

It has been already described in the previous section that the LSRP is similar to the LSC in allowing a periphery for each of its layers. The uppermost PER node (PER-RP) can be occupied by non-restrictive modifiers, as are appositional phrases and some relative clauses:

PER-RP-> RP || MP || CL

(29) The king, aware of the lack of enthusiasm of his subjects….

The core periphery rule is:

PER-CRP->PP || MP || PP PP.

It accounts for setting PPs and MPs, as in:

(30) The election of a new President *in Las Cortes yesterday*

And the rule for the innermost nuclear periphery is

PER-NRP-> MP MP MP || CL

as this node can be saturated either by restrictive MPs and relative clauses:

(31) A *long peaceful rewarding* holiday *which brought peace to our lives*

Note that some demonstratives, quantifiers and numerals can also appear as NUC-RP in an RP with a layered structure. This happens when they take partitive prepositional arguments, as in:

(32) *Those of you* who can play chess

(33) *Some* of my friends

(34) *One* of these days

(35) The *last* of his requests

Such an option is expressed by the rule proposed for NUC-RP:

NUC-RP-> N || ADJ || ADJ PER-NRP || **PROD || PROQ || NUMC || |NUMO ||…**

The rule also contemplates the various possibilities that may arise from the combination of the head (and, if relevant, the nuclear peripheral elements) with a range of quantifier and numeral postdeterminers, as in the following examples:

(36) the first$_{\text{NUMO}}$ two$_{\text{NUMC}}$ exhausted$_{\text{PER-NRP}}$ runners$_{\text{N}}$ who managed to arrive at the line $_{\text{PER-NRP}}$ (NUMO NUMC PER-NRP N PER-NRP)

(37) the three$_{\text{NUMC}}$ last$_{\text{NUMO}}$ coins$_{\text{N}}$ that were left in the coffin$_{\text{PER-NRP}}$ (NUMC NUMO N PER-NRP)

(38) the many$_{\text{DETQ}}$ brightly coloured$_{\text{PER-NRP}}$ window panes$_{\text{N}}$ (DETQ PER-NRP N)

A crucial element in the LSRP is the so-called RP Initial Position (RPIP). The rule to account for its configurational possibilities is:

RPIP -> ART || DETP || DETD || RP || MP || PROD || PROP

The RPIP marks the definiteness of the RP; therefore it hosts the central determiners, including here the articles, demonstratives and possessives[13]; central determiners may in turn be modified by a group of partitive predeterminers like *all, both, double, half, twice, many (a, such (a), what (a)*. RPIP is also the position for genitive RPs and MPs (*Yesterday's news*), which are also intrinsically definite. Note that if a possessed RP is indefinite, it will appear as a peripheral post nuclear PP within the Core$_{\text{MP}}$ (as in *a son of Peter's*).

The following rule in this subset explains the internal structure of CORE-RP: Apart from the Nucleus, and the peripheral elements it includes the possibility of having one, two or three arguments. This seems to be the maximum of arguments allowed and it happens with some derived nominals as NUC-RP:

(39) The donation of 100.000 dollars to the asylum by an anonymous benefactor

ARGs in RPs can be PP, as in the example above, or CL:

[13] The difference between possessives and demonstratives that can occur in RPIPs and those that are RPs by themselves is captured in ARTEMIS by considering them members of different POS. Whereas the former are considered subtypes of the DET category: DETD, DETD, etc, the latter are members of the PRO set: PROP, PROD. Etc.

(40) His suggestion that they should not let her take control

(41) The desire to escape from prison

In the following example we can see both types occurring together within the same CORE-RP:

(42) The convention *of one Philippine tribe that no man can keep a secret* (example from Downing and Locke (1992: 463).

The proposal for new syntactic categories in Van Valin (2008) also affected AdjPs and AdvPs, which are now grouped into a single functional node, the Modifier Phrase. The last subset of rules above accounts for the internal configuration of this type of constituent:

(1.c) Modifier Phrases
MP -> PER-MP CORE-MP || CORE-MP
CORE-MP-> NUC-MP || NUC-MP ARG-MP
NUC-MP - > ADJ || ADV || N || RP || CL || S
ARG-MP -> PP || CL
PER-MP -> MP

Typically, the MP modifying a noun has either and ADj or an N as nucleus (as in *An impressive still-life* and a *meat knife,* and the MP modifying a verbal nucleus or a core have an ADV nucleus (as in *He progresses slowly*), but Van Valin (2008: 172) explains that other categories can occupy such a position, as revealed by the following examples from Lieber (1992)

(43) The *Charles and Di* syndrome is no longer relevant

and Everett (2006):

(44) The *God is dead* philosophers are mostly dead

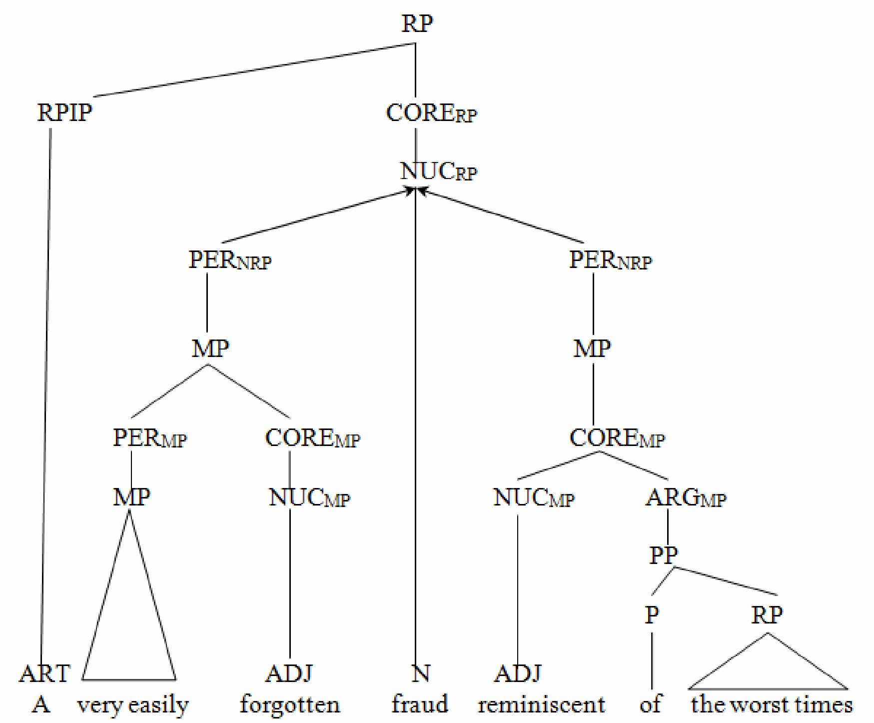
(45) My grandson likes to give me the *who's the boss now, silly old grandpa* wink frequently

All these possibilities are expressed in the following rule:

NUC-MP - > ADJ || ADV || N || CL || S.

The other rules spell out the layered structure of MPs (LSMP) in accordance to what was described in section 3. The following figure illustrates the LSMPs of *very easily* *forgotten* and *reminiscent of* the *worst times* in the RP *a very easily forgotten fraud reminiscent of the worst times:*

(46)



The rule for the core of MPs:

CORE-MP-> NUC-MP || NUC-MP ARG-MP

captures the fact that in English predicative adjectives can have arguments, as happens in *My little brother is afraid of the dark*; on the contrary, attributive adjectives cannot take arguments in English (**My afraid of the dark little brother*). In fact, in Van Valin and LaPolla (1997: 69) this was taken as an argument to defend that they did not have a layered structure. Van Valin (2008: 172), based on Matasović (2001), corrects this view as there are languages which do have attributive adjectives taking an argument. This has led us to introduce a disjunctive option in the rule for CORE-MP, even though there are no core arguments for the attributive NUC_{MP} in English[14].

---

[14] Since ARTEMIS is primarily a left-to-right and bottom-up with top-down prediction parser there is no danger of a wrong assignment of attributive NUC + argument MP.

Note that the rules as stated above are not complete since the nodes do not have their corresponding AVMs. This has been done to ease the explanation about the syntactic configuration of RPs, MPs and PPs. Let us turn now to consider the AVMs for the relevant constituents within such rules; some of them were already included in the original AVMs list within ARTEMIS, and have only been subject to certain adjustments; the rest are to be introduced as new. Both the modifications and the new AVMs are marked in boldface:

(47)

<Category Type="**ART**">
<Attribute ID="Count" />
<Attribute ID="Def" />
<Attribute ID="Num" />
</Category>

<Category Type="CORE-RP">
<Attribute ID="Count" />
<Attribute ID="Num" />
<Attribute ID="Pol" />
</Category>

<Category Type="DETD">
<Attribute ID="Def" />
<Attribute ID="Num" />
<Attribute ID="Prox" />
</Category>

<Category Type="DETP">
<Attribute ID="Def" />
<Attribute ID="Num" />
<Attribute ID="Per" />
</Category>

<Category Type="DETQ">

</Category>

<Category Type="NUC-RP">
<Attribute ID="Count" />
<Attribute ID="Num" />
</Category>

<Category Type="NUMC">
<Attribute ID="Num" />
</Category>

<Category Type="NUMO">
<Attribute ID="Num" />
</Category>

<Category Type="PART">
<Attribute ID="Num" />
</Category>

<Category Type="PROP">
<Attribute ID="Num" />
<Attribute ID="Per" />
</Category>

<Category Type="RP">
<Attribute ID="Case" />

```
    <Attribute ID="Count" />
    <Attribute ID="Num" />
    <Attribute ID="Pol" />
    <Attribute ID="Quant" />
</Category>
```

```
    <Attribute ID="Count" />
    <Attribute ID="Def" />
    <Attribute ID="Num" />
    <Attribute ID="Prox" />
</Category>
```

```
<Category Type="NOUN">
    <Attribute ID="Case" />
    <Attribute ID="Concept" />
    <Attribute ID="Count" />
    <Attribute ID="Num" />
```

```
<Category Type="RPIP">
    <Attribute ID="Def" />
    <Attribute ID="Num" />
    <Attribute ID="Prox" />
</Category>
```

The only new Attributes that have been introduced in the above AVMS are “Pol”, which refers to the ‘Polarity’ of some quantifiers like any, some, no and its compound derivatives, “Quant” which stands for “quantification” and is also an attribute of quantifier words, and “Case” which is a cancellable attribute for RPs and percolates up from nominal heads. Quantification can be relative (few, many, little) or absolute (all, no), and positive (many) or negative (few) in terms of quantity. Quantifiers interact in very interesting ways with sentential negation and no has a especially relevant role as the exponent of RP negation in English (no books). The corresponding AVMs for these Attributes will be:

(48)

```
<Attribute ID="Pol" obl="+" num="1">
   <Value Tag="assertive">a</Value>
   <ValueTag="non-assertive">na
         </Value>
    <Value Tag="negative">n</Value>
   </Attribute>

<Attribute ID="Case" obl="*" num="1">
   <Value Tag="genitive ">gen
         </Value>
    </Attribute>
```

```
<Attribute ID="Quant" obl="+" num="1">
   <ValueTag="relative positive">rp
         </Value>
   <ValueTag="relative negative">rn
         </Value>
   <ValueTag="absolute positive">ap
         </Value>
    <Value Tag="absolute negative">an
         </Value>
   </Attribute>
```

In order to illustrate how AVMS and their attributes should be encoded in the rules in ARTEMIS, (49) shows the final aspect of the rule for RPIP constituents:

(49) RPIP [Def = ?def, Num = ?num, Prox = ? prox] -> PART [Num= pl | sg] ART [Count = c | u , Def= d | i, Num= pl | sg]|| PART [Num= pl | sg] DETP [Def= d, Num = pl | sg, Per= 1 | 2 | 3] ||PART [Num= pl | sg] DETD [Def= d, Num = pl | sg, Prox= n | y]|| ART [Count = c | u , Def= d | i, Num= pl | sg] || DETP [Def= d, Num = pl | sg, Per= 1 | 2 | 3]||DETD [Def= d, Num = pl | sg, Prox= n | y] || RP [Case = ?case, Count= ? cnt, Def= ?def, Num= ?num, Prox= ?prox]|| MP [Degree = ? deg, Form= ?form, Pst= ?pst]

The same must be done with the rest of the rules for XPs proposed in this paper.

5. Conclusion

As stated in the introduction, the analysis in this paper aimed at contributing to the development of the GDE within a functionally based parsing prototype, ARTEMIS. In order to do so, it has been necessary to previously offer an updated description of phrasal units in Role and Reference Grammar. Consequently, this study also contributes to the computational implementation of this linguistic theory. There are, however, several pending issues to achieve these goals, among which the following must be considered: (i) the revision of the lexical rules associated to the functional units that are encoded as POS tags in the repository for lexical items within ARTEMIS. Several of these tags are used in the rules proposed for XPs in this paper; (ii) AVMs need significant revision, subject to the development of further analysis both from within and without phrasal constituents; and (iii) the prototype rules for the different layers that make up the so-called LSC in RRG needs thorough refinement.

References

Bird, S., E. Loper, and E. Klein. 2009. *Natural Language Processing with Python*. Beijing, Cambridge, Farnham, Köln, Sebastopol, Tokyo: O'Reilly Media Inc.

Boas, H.C and I.A. Sag. 2012. *Sign-Based Construction Grammar*. Stanford, Cal.: CSLI Publications.

Díaz Galán, A. and M.C. Fumero Pérez. 2015. "Developing parsing rules within ARTEMIS: the case of DO auxiliary insertion". In C. Periñán-Pascual and E. Mestre i Mestre (eds) 2015 *Understanding Meaning and Knowledge Representation: From Theoretical and Cognitive Linguistics to Natural Language Processing* (in press). Newcastle: Cambridge Scholars Press.

Diedrichsen, E. 2014. "A Role and Reference Grammar Parser for German". In B. Nolan and C. Periñán-Pascual (eds) 2014 *Language Processing and Grammars*. (pp. 105-142). Amsterdam/Philadelphia: John Benjamins.

Downing, A., and P. Locke. 1992. *A University Course in English Grammar*. New York, London, Toronto, Sydney, Tokyo, Singapore: Prentice Hall.

Everett, D. 2006. "Wari' intentional state constructions". Ms. Available ling.auf.net/lingbuzz/000438/current.pdf

Guest, E. 2009. "Parsing using the Role and Reference Grammar paradigm". <http://repositoryintralibrary.leedsmet.ac.uk/open_virtual_file_path/i42n47491t/Parsing Using the Role and Reference Grammar Paradigm.pdf> (21 September 2012).

Lieber, R. 1992. *Deconstructing Morphology: Word Formation in Syntactic Theory*. Chicago, IL: University of Chicago Press.

Mairal Usón, R. and F.J. Ruíz de Mendoza. 2008. "New challenges for lexical representation within the Lexical-Constructional Model (LCM)". *Revista Canaria de Estudios Ingleses*, 57: 137-158

Mairal Usón, R. and F.J. Ruíz de Mendoza. 2009. "Levels of description and explanation in meaning construction. In C. Butler and J. Martín Arista (eds) 2009 *Deconstructing Constructions*. (pp. 153-200) Amsterdam: John Benjamins.

Matasović, R. 2001. "Adjective Phrases". Ms. Available http://wings.buffalo.edu/linguistics//people/faculty/vanvalin/rrg/Matasovic.pdf

Nunes, M.L. 1993. "Argument linking in English derived nominals". In Van Valin, R.D. Jr. (ed) 1993 *Advances in Role and reference Grammar* (pp. 375-432). Amsterdam and Philadelphia: John Benjamins.

Pavey, E.L. 2010. *The Structure of English. An Introduction to Grammatical Analysis*. Cambridge: Cambridge University Press.

Periñán-Pascual, C. 2013. "Towards a Model of Constructional Meaning for natural Language Understanding." In B. Nolan and E. Diedrichsen (eds) 2013 *Linking Constructions into Functional Linguistics: The role of constructions in grammar* (pp. 205-230). Amsterdam: John Benjamins, 2013.

Periñán-Pascual, C. and F. Arcas-Túnez. 2010. "The Architecture of FungramKB." *Proceedings of the 7th International Conference on Language Resources and Evaluation* (pp. 2667-2674). Malta: European Language Resources Association.

Periñán-Pascual, C and F. Arcas-Túnez. 2014. "The implementation of the CLS constructor in ARTEMIS". In B. Nolan and C. Periñán-Pascual (eds) 2014 *Language Processing and Grammars* (pp. 165-196). Amsterdam: John Benjamins.

Periñán-Pascual, C. and R. Mairal Usón. 2009. "Bringing Role and Reference Grammar to natural language understanding". *Procesamiento del Lenguaje Natural*, núm. 43: 265-273.

Periñán-Pascual, C. and Mairal Usón. 2011 "The COHERENT Methodology in FunGramKB". *Onomazein*. 24: 13-33.

Ruiz de Mendoza Ibáñez, F. and R. Mairal Usón. 2008. "Levels of Description and Constraining Factors in Meaning Construction: an Introduction to the Lexical Constructional Model." *Folia Linguistica* 42-2: 355-400.

Sag, I. A., T.Wasow and E. M. Bender. 2003. *Syntactic Theory: Formal Introduction*. Stanford: CSLI Pulbications..

Salem, Yasser, Hensman, Arnold and Nolan, Brian. 2008. Towards Arabic to English machine translation. *ITB Journal* 17: 20–31.

Van Valin, R. D. Jr. 2005. *Exploring the Syntax-Semantics Interface*. Cambridge: Cambridge University Press.

Van Valin. 2008. “RPs and the nature of lexical and syntactic categories in Role an Reference Grammar”. In R.D. Van Valin Jr. (ed) 2008 *Investigations of the Syntax-Semantics-Pragmatics Interface* (pp. 161-178). Amsterdam/Philadelphia: John Benjamins.

Van Valin and R. LaPolla. 1997: *Syntax: Structure, Meaning and Function*, Cambridge: Cambridge University Press.

Van Valin and R. Mairal. 2014. “Interfacing the lexicon and an ontology in a linking system”. In M. A. Gómez, F. Ruiz de Mendoza and F. Gonzálvez-García (eds) 2014 *Theory and practice in functional-cognitive space* (pp. 205-228). Amsterdam: John Benjamins.

Rijkhoff, J. 1990. “Explaining word order in the noun phrase”. *Linguistics* 28: 5-42.

Rijkhoff, J. 1992. *The noun phrase: A typological study of its form and structure*. Unpubished PhD Dissertation. University of Amsterdam.

Appendix: List of Abbreviations

Constituents and Parts of Speech in Parsing Rules

ADJ	Adjective
ADV	Adverb
ARG	Argument
ARG-MP	Argument of MP
ARG-RP	Argument of RP
ART	Article
AUX	Auxiliary verb
CL	Clause
CONSTR-L1	Level 1 Construction
CORE-MP	MP Core
CORE-PP	PP Core
CORE-RP	RP Core

DETD	Determiner (demonstrative)
DETP	Determiner (possessive)
DETQ	Determiner (quantifier)
LDP	Left detached Position
MP	Modifier Phrase
N	Noun
NOUX	Proper noun
NUC	Nucleus
NUC-A	Adjunct Nucleus
NUC-MP	MP Nucleus
NUC-P	Preposition Nucleus
NUC-RP	RP Nucleus
NUC-S	Secondary Nucleus
NUMC	Numeral (cardinal)
NUMO	Numeral (ordinal)
PART	Pronoun (partitive)
PER	Periphery
PER-CRP	Periphery of RP Core
PER-MP	MP Periphery
PER-NRP	Periphery of RP Nucleus
PER-PP	PP Periphery
PER-RP	RP Periphery
PP	Prepositional Phrase
Pre-CL1	Pre L1 Construction
PRED	Predicate
PRED-A	Adjunct Predicate
PRED-S	Secondary Predicate
PRO	Pronoun
PROD	Pronoun (demonstrative)
PROP	Pronoun (Possessive)
PROQ	Pronoun (quantifier)
RDP	Right Detached Position
RP	Referential Phrase
RPIP	RP Initial Position
S	Sentence

Attributes in AVMs

Var	Variable
Illoc	Illocutionary force
INCH	Inchoative construction
INST	Instrument Construction
KER1	Kernel 1 Construction
KER2	Kernel 2 Construction
MIDD	Middle Construction
CMOT	Caused Motion Construction
Num	Number
Prep	Preposition
Count	Countability
Def	Definiteness
Num	Number
Pol	Polarity
Prox	Proximity
Per	Person
Quant	Quantification

Received: February 10, 2015

Accepted: February 26, 2016

Published: February 29, 2016

Updated: March 2, 1016