Parsing WH-interrogative sentences within ARTEMIS[1]

Parseado de las interrogativas con formas WH- en ARTEMIS

MARÍA AUXILIADORA MARTÍN-DÍAZ
UNIVERSITY OF LA LAGUNA
INSTITUTE OF LINGUISTICS ANDRÉS BELLO

The linguistic foundation of Functional Grammar Knowledge Base (FunGramKB) is inspired on robust linguistic theories like Role and Reference Grammar (RRG, Van Valin, 2005) and the Lexical Constructional Model (LCM) (Mairal-Usón & Ruiz de Mendoza Ibáñez, 2008). Accordingly, its lexical and grammatical levels will allow the system to capture syntactic-semantic generalizations which can provide explanations and predictions of language phenomena (Periñán-Pascual & Arcas-Túnez, 2010: 2671). ARTEMIS (Automatically Representing Text Meaning via an Interlingua-Based System) comes into play as a prototype which exploits FunGramKB to pursue the simulation of natural language understanding. To guarantee a correct parsing, this application relies on a set of production rules where pertinent morphosyntactic information must be expediently encoded via AVMs, feature-bearing structures which can effectively constrain this process (Periñán-Pascual, 2013a: 223). Departing from Martín-Díaz (2017)'s account of Y/N-interrogatives, this paper aims at proposing the rules and AVMs necessary for the correct parsing of WH-interrogatives within ARTEMIS.

Keywords: *WH-Questions; Role and Reference Grammar; ARTEMIS*

El fundamento lingüístico de Functional Grammar Knowledge Base (FunGramKB) se inspira en teorías lingüísticas como la Gramática del Papel y la Referencia (RRG) (Van Valin, 2005) y el Modelo Léxico-Construccional (LCM) (Mairal-Usón y Ruiz de Mendoza Ibáñez, 2008). Sus niveles léxico y gramatical permiten al sistema capturar generalizaciones sintáctico-semánticas con las que explicar y predecir fenómenos lingüísticos (Periñán-Pascual y Arcas-Túnez, 2010: 2671). El objetivo de ARTEMIS (Automatically Representing Text Meaning via an Interlingua-Based System) es la simulación de la comprensión del lenguaje natural y para ello utiliza FunGramKB. Este prototipo garantiza el correcto parseado con unas reglas de producción donde la información morfosintáctica pertinente debe codificarse mediante Matrices Atributo Valor (MAVs) que puedan restringir dicho proceso (Periñán-Pascual, 2013a: 223). Partiendo del análisis de Martín-Díaz (2017) para las interrogativas polares, este trabajo propone las reglas y MAVs necesarias para un parseado efectivo de las interrogativas inglesas con formas WH- en ARTEMIS.

Palabras clave: *Preguntas con formas WH- inicial; Gramática del Papel y la Referencia; ARTEMIS*

---

1. INTRODUCTION

Functional Grammar Knowledge Base (FunGramKB) is linguistically grounded in sound language theories like Role and Reference Grammar (RRG) (Van Valin, 2005) and the Lexical Constructional Model (LCM) (Mairal-Usón & Ruiz de Mendoza Ibáñez, 2008; Ruiz de Mendoza Ibáñez & Mairal-Usón, 2008). The close interrelation between these grammatical theories and FungramKB is evident in the design of ARTEMIS (Automatically Representing Text Meaning via an Interlingua-Based System"), an NLP system whose objective is the simulation of natural language understanding. ARTEMIS becomes a development of the syntax-to-semantics linking algorithm proposed in RRG and, by capturing syntactic-semantic generalizations, it is able to provide both explanations and predictions of language phenomena (Periñán-Pascual & Arcas Túnez, 2010; Periñán-Pascual, 2013a).

A summary of the tenets of importance for this analysis within RRG as well as the fundamentals followed by FunGramKB and ARTEMIS will be offered in this introduction. Section 2 will be devoted to considering a classical description of WH-Questions as the one propounded by Quirk, Greenbaum, Leech and Svartvik (1985: 806-838), together with a brief account of the English WH-words used in such structures. Section 3 will deal with the analysis of WH-Questions within RRG and Section 4 with the implementation of questions within ARTEMIS. Section 5, in turn, will specifically address the representation of the necessary production rules for an accurate parsing of English WH-interrogative structures within our NLP prototype, as well as a discussion of the nuts and bolts of devising an AVM (Attribute-Value Matrixes) for each of the WH-forms taking part in this type of interrogatives.[2] The conclusion in Section 6 will finally display the main achievements attained in this paper.

As a knowledge engineering tool specially designed for simulating natural language comprehension, FunGramKB can be described as an NLP artefact whose central spine grows along two independent yet interrelated modules. On one side we have the linguistic module, a knowledge repository integrated by two language specific constituents, the lexical and the grammatical levels. A knowledge repository comprising an ontology, a cognicon and an onomasticon occupies the other side, being thus a more abstract and universal module (Periñán-Pascual & Mairal-Usón, 2011).

ARTEMIS comes into play at this point as the NLP system which will use FunGramKB as a knowledge base for the automatic representation of natural language sentences. This prototype, theoretically grounded on a linguistic model like RRG has also been greatly influenced by the LCM, a cognitive-functional model. Although the original adoption of functional aspects like the syntax-semantics linking algorithm, the Logical Structures (LSs) and the Layered Structure of the Clause (LSC) proved useful for text meaning representation, it needed to be enriched with a deeper semantically oriented theory like the LCM which would provide FunGramKB with an enhanced semantic capacity (Periñán-Pascual, 2013a). ARTEMIS then involves a parsing process where the grammatical units and nodes in the LSC are processed in order to generate Conceptual Logical Structures (CLSs) where variables directly relate to the conceptual information stored in FunGramKB (Periñán-Pascual & Arcas

---

[2] See section 4 below for a detailed description of these AVMs, as well as a description of the corresponding lexical rules.

Túnez, 2014).[3] This conceptual shift from LSs to CLSs facilitates the mapping into a COREL scheme, a formalized structure whose metalanguage is more adequate from a computational viewpoint and can be more easily interpreted by an automated reasoner.[4]

The following example taken from (Periñán-Pascual & Arcas Túnez, 2014) can help us to illustrate this conceptual shift from LSs to CLSs:

(1) Peter broke the glass.

RRG Logical structure:

(2) $<_{IF}^{DEC} <_{TNS}^{PAST} <_{ASP}^{PERF}$ <[**do'** (Peter, Ø)] CAUSE [BECOME **broken'** (glass)]>>>>

FunGramKB CLS:

(3) $<_{IF}^{DEC} <_{TNS}^{PAST} <_{ASP}^{PERF} <_{CONSTR-L1}^{KER2} <[_{AKT}^{CACC}$ [+BREAK_00 (%PETER_00-Theme, $GLASS_00-Referent)]]>>>>

As can be seen, the RRG logical structure corresponds to a causative accomplishment in which variables are instantiated by predicates like *glass* and *Peter*, or primitives like **broken'**. In the CLS, on the other hand, these variables are saturated by ontological concepts like $GLASS_00 and %PETER_00 to which a thematic role in the corresponding thematic frame of +BREAK_00 is assigned.[5] That is, %PETER_00 becomes the Theme and $GLASS_00 the Referent of the event +BREAK_00. The cognitive situation described in this CLS also includes a CONSTR-L1 operator, which marks a Kernel 2 argumental construction, and an AKT operator which characterizes its *Aktionsart* as a causative accomplishment (see Periñán-Pascual, 2013a, and Periñán-Pascual & Arcas-Túnez, 2014 for a detailed description of how this CONSTR-L1 operator will consequently have a corresponding node in the enhanced LSC).

Even though the semantic representation has been conceptually enhanced in the CLS, it still needs some computational refining if we want to reach a deep level of comprehension in our syntactic parser. This refinement is achieved in ARTEMIS by translating the CLS into a COREL scheme, as the following example retrieved from Periñán-Pascual and Arcas-Túnez (2014), and Cortés-Rodríguez and Mairal-Usón (2016) shows:

(4) +(e1: +DAMAGE_00 (x1: %PETER_00)Theme (x2: $GLASS_00)Referent (f1: (e2:+SPLIT_00 (x1)Theme (x2)Referent))Result)
‘Peter damaged the glass into pieces’

In ARTEMIS, the COREL-scheme Builder and the CLS Constructor, together with a Grammar Development Environment (GDE) constitute the central components of a modular system. The GDE module stores a catalogue of Attribute-Value Matrixes (AVMs) and a set of production rules (lexical, constructional and syntactic) necessary to build a feature-based grammar able to constrain the parsing process (Periñán-Pascual, 2013a). A feature unification

---

[3] A CLS is an enriched version of an RRG's Logical Structure (LS).
[4] COREL (COnceptual REpresentation Language).
[5] FunGramKB's ontology distinguishes three levels of knowledge: a core level where pivotal basic concepts are headed by '+'; a higher universal level where metaconcepts (preceded by a hash-tag #) "can facilitate ontological interoperatibility"; and a lower particular level where terminal concepts identified with a $-symbol "can grant immediate applicability" (Periñán-Pascual, 2013b).

process relates these two constructs of the GDE and demands the use of such AVMS as non-atomic meaning-bearing devices that come to replace RRG's operator projection. Besides, as Cortés-Rodríguez & Mairal-Usón (2016: 102) claim "unification processes require for grammatical features to be encoded not only in the layer which they modify, but in all the nodes dominated by such a layer down to the lexical token which is the formal expression of the operator concerned".

Feature Unification involves the percolation from such a lexical token to the node over which the Operator has scope. In the case of the AVM for the interrogative CORE, the attribute *Illoc* (for illocutionary force) always appears encoded in the first constituent of the CORE (the first auxiliary verb or the lexical predicate if there are no auxiliaries), even though the percolation process[6] will finally occur in the layer over which this operator has scope, the Clause level. See Fig. 1[7] below for an illustration of this unification process in a YNQ.[8]

We deal here with WH-interrogative structures and, apart from providing their corresponding syntactic rules for a correct parsing to be done in ARTEMIS, which includes highlighting the importance of the presence of WH-forms in RRG's PreCore Slot, we also have to encode the vital information conveyed by RRG grammatical operators like illocutionary force, modality and/or status.[9] Likewise, AVMs for the different WH-forms presenting these structures should be also accounted since the features lodged in them will constrain the phrase structure rules of our interrogatives, as will be also shown in Section 4 below.
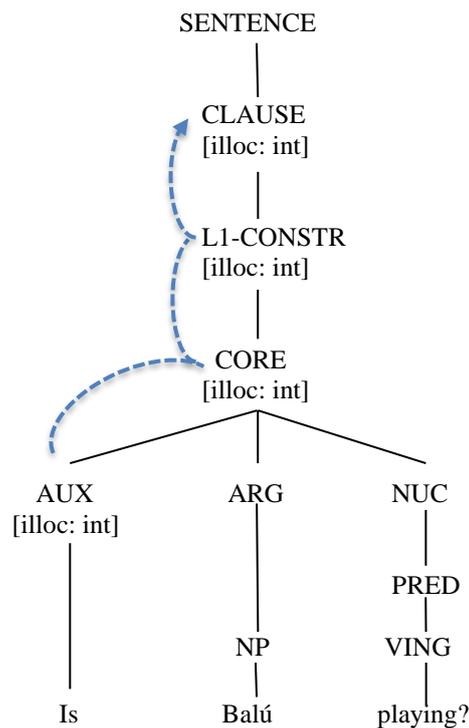
SENTENCE

|

CLAUSE
[illoc: int]

|

L1-CONSTR
[illoc: int]

|

CORE
[illoc: int]

AUX   ARG   NUC
[illoc: int]

| | |
             NP   PRED
                          |
             | VING
                          |
Is   Balú   playing?

*Fig. 1: **Feature Unification Path of Illocutionary Force Attribute***

---

[6] Or Feature-Unification Path in Cortés-Rodríguez and Mairal-Usón (2016)

[7] This figure has been adapted from Fig. 13 in Cortés-Rodríguez and Mairal-Usón (2016). In it, the L1-CONSTR node has been incorporated to the LSC.

[8] Abbreviation for Yes-No questions. See Martín-Díaz (2017) for a full treatment of these interrogative structures in ARTEMIS.

[9] See section 4 below for a detailed description of this important slot in these WH-interrogative structures.

2. WH-INTERROGATIVES: WH-WORDS, AUXILIARIES, MODALS AND OPERATORS

The category of WH-Questions stands out as one of the largest categories of interrogative structures in English. Their analysis is based on a traditional classification of questions proposed by Quirk et al. (1985: 806-825).[10] Out of the three fundamental aspects highlighted by these authors in the formation of WH-Questions, two (the inversion of the syntactic order and the use of primary/modal auxiliaries) have already been described in Martín-Díaz (2017), where on accounting for the parsing of Yes-No Questions within ARTEMIS, a review of this type of interrogatives is offered.

Nevertheless, an important caveat is due with respect to this prototypical interrogative word order. According to Quirk et al., in WH-interrogative clauses the Subject-Operator (S-O) inversion occurs except where the WH-word is itself the subject (1985: 725). This fact will necessarily have to have an impact on the design of the corresponding production rules for WH-interrogatives in the GDE.

(5) *What have* (O) *you* (S) *seen today?*
(6) *What* (S) *has* (O) *kept you so long?*

The third fundamental aspect underlined by Quirk et al. (1985) in relation to these interrogatives involves the analysis of the WH-element, a segment which a prototype like ARTEMIS will certainly have to control in order to produce an effective parsing. WH-structures include WH-words like *which*, *when*, *why*, *where*, *what*, *who*, *whose*, *whom*, and *how*. Through their use in initial position we can ask for the identification of the subject, object, complement, or adverbial of a sentence (Quirk et al., 1985: 77). Therefore the grammatical category of these WH-words can be of a very different kind:[11]

1) Pronoun (*Who are you?*)
2) Time adverb (*When do they ....?*)
3) Modifying adverb (*How old are you?*)
4) Determiner (*Which cup is yours?*)
5) Adjective (*How do you feel?*)

3. RRG'S INTERROGATIVES

Since FunGramKB and therefore ARTEMIS are based on RRG as a fundamental grammatical model, some matters on how this grammar handles WH-Questions in general and WH-forms (word or phrases) need considering.

In RRG, the PreCore slot (PrCS) in the constituent analysis of sentences is the place typically occupied by question-words in languages in which they do not appear in situ (see the tree in Fig. 2 below for an illustration of the interrogative *What did you buy in that shop?*).

---

[10] For details, see Quirk et al. (1985: 825 ff.).
[11] As we will see below, considerations like these imply introducing modifications in ARTEMIS, basically because new tags and new tokens will need to be registered in the POS section (Parts of Speech) and new AVMs will have to be designed for new categories and attributes.
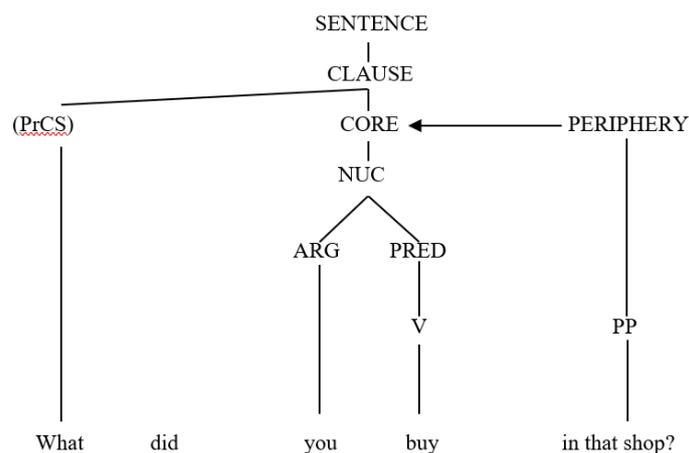
```
                        SENTENCE
                           |
                        CLAUSE
        _____/        \
   (PrCS)                  CORE  ◄──────── PERIPHERY
                            |
                           NUC
                          /    \
                       ARG     PRED
                                |
                                V                    PP


   What      did      you     buy          in that shop?
```

Fig. 2

Perhaps the most important language specific qualification of Van Valin's syntactic template selection principle has to do with interrogative structures like these where the presence of a WH-syntactic argument in the PrCS implies a reduction in the number of core slots.[12] This principle explains the diminution of CORE arguments in 7b (with a semantic argument in the PrCS) and not in 7c (whose PrCS is not an argument of the CORE but an adjunct) below, and also in the most basic frames for WH-Questions in ARTEMIS, as will be seen in the following section:

(7)

      a) *She* (CORE ARG) *wrote me* (CORE ARG) *a poem* (CORE ARG) *yesterday* (ADJ) → 3 CORE ARGUMENTS.

      b) *Who* (ARG in PrCS) *wrote me* (CORE ARG) *a poem* (CORE ARG) *yesterday* (ADJ)? → 2 CORE ARGUMENTS.

      c) *When* (ADJ in PrCS) *did she* (CORE ARG) *write me* (CORE ARG) *a poem* (CORE ARG)? 3 CORE ARGUMENTS.

Besides, in the interpretation of RRG syntactic categories Van Valin (2005) introduces some important variations to the ones originally described in both Van Valin and LaPolla (1997) and Van Valin (2005). He propounds Referential Phrases (RPs) and Modifier Phrases (MPs) as two types of constituents more functionally and typologically valid. In keeping with this proposal Cortés-Rodríguez (2016b) discusses the rules for the syntactic parsing of RRG's phrasal constituents and offers a description of how the layered structure assigned to NPs can now adapt to what we may call the Layered Structure of Referential Phrases (LSRP).

We raise this issue at this point because in line with the previous proposals we will analyse the fronted-constituent (WH-words or WH-phrases) of our WHQs as part of the above mentioned LSRP and as arguments in the PrCS, an extracore slot which as explained later on in section 4 has been relabeled PreC-L1 position.

4. THE IMPLEMENTATION OF ENGLISH WH-QUESTIONS IN ARTEMIS

In the unification process that relates the two constituents in the GDE, ARTEMIS will need,

---

[12] "The occurrence of a syntactic argument in the PrCS reduces the number of core slots by 1" (Van Valin & LaPolla, 1997: 174).

as explained in the previous section, the syntactic rules indispensable for the parsing of WHQs and, on the other hand, the spelling out of the relevant AVMs that will guarantee their correct parsing[13].

Basically, WHQs inherit from YNQs the codification for their syntactic rules in the GDE (Martín-Díaz, 2017). However, the introduction of a CONSTR-L1 node in Periñán-Pascual and Arcas-Túnez (2014: 171-175) in order to distinguish kernel from non-kernel constructions has implied the redefinition of RRG's PrCS as a Preconstruction-L1 position (PreC-L1 node in Cortés Rodríguez, 2016b). It is important to highlight this circumstance for two reasons: firstly, because of the consequences the presence of this WH-fronted component can have in the subsequent CORE-frame (see RRG's syntactic template selection principle mentioned in the previous section); secondly, because it is precisely in this PreC-L1 position where the WH-component (word or phrase) of our WHQs is located (see Fig. 3 below for an illustration).

Cortés-Rodríguez (2016b) becomes crucial for a correct parsing of English WHQs within ARTEMIS because on discussing about the rules for the syntactic parsing of RRG's phrasal constituents he expands the layered structure of PP (Prepositional Phrases), RPs (Referential Phrases) and MP (Modifier Phrases). As constituents of the PreC-L1 slot of our interrogative structures, we also need a correct spelling out of the syntactic rules for our WH-words or phrases in the GDE. To begin with, such rules prompt for certain adjustments in ARTEMIS, like for example:

1) The introduction of a new token for the category PROI (interrogative pronoun) in the Parts of Speech (POS). This tag includes *what, which, who, whom*, but lacks the token *whose*.
2) The category ADVI (interrogative adverb) is empty and the tokens *where, when, how* and *why* need to be included.
3) A tag for a new category DETI (interrogative determiner) needs to be registered in order to include tokens like *whose, what* and *which*.

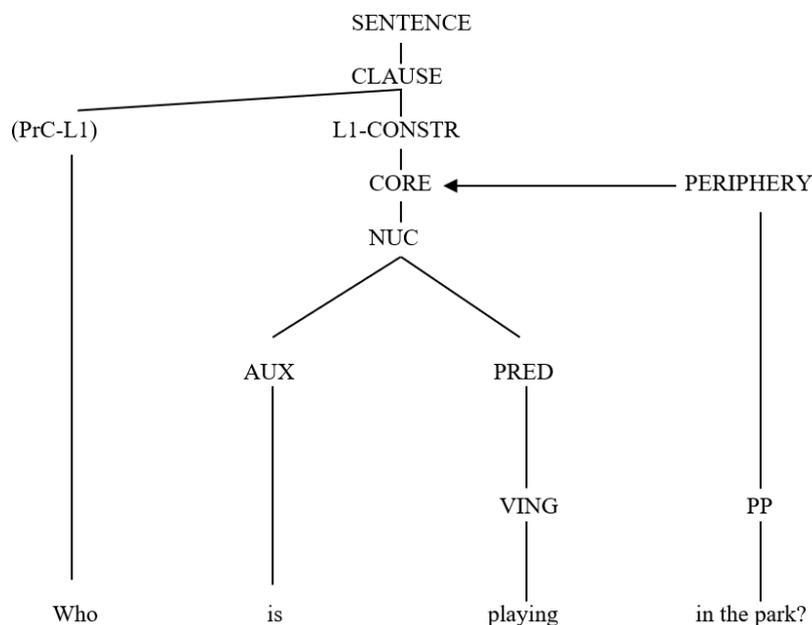SENTENCE
|
CLAUSE

(PrC-L1)   L1-CONSTR
|
CORE ◄———————————— PERIPHERY
|
NUC

AUX   PRED

VING   PP

Who   is   playing   in the park?

Fig. 3

---

[13] The tag used in the initial design of the GDE for a WH-Question is SBARQ, but for the sake of clarity here we propose to change it for WHQ. Likewise, we have already proposed to replace with YNQ the SQ-tag used in ARTEMIS for YES/NO questions (Martín-Díaz, 2017).

Subsequently, new AVMs for these new categories must be codified and some new attributes (i.e., case and animacy) and values (i.e., genitive, objective and subjective; and nonpersonal and personal) will have to be introduced too:

New AVMs for Categories and Attributes:
(8) <Category type="PROI" >
     <Attribute ID>="Animacy"/>
     <Attribute ID>="Case"/>
     <Attribute ID>="Illoc"/>
     <Attribute ID>="Reference"/>
     <Attribute ID>="Role"/>
     <Category>

(9) <Category type="ADVI" >
     <Attribute ID>="Illoc"/>
     <Attribute ID>="Role"/>
     <Category>

(10) <Category type="DETI" >
      <Attribute ID>="Case"/>
     <Attribute ID>="Illoc"/>
     <Attribute ID>="Reference"/>
     <Attribute ID>="Role"/>
     <Category>

(11) <Attribute ID="Case" obl="*" num="1">
      <Value Tag="genitive">gen
      <Value Tag="objective">obj
      <Value Tag="subjective">subj
      </Value>
      </Attribute>

(12) <Attribute ID="animacy" obl="*" num="1">
      <Value Tag="nonpersonal">nonp
      <Value Tag="personal">personal
      </Value>
      </Attribute>

Lexical rules:[14]

(13) PROI's Lexical rules
     WHAT [Animacy=personal ∣ nonp, Case=null, Illoc:int, reference=indef, role=Referent ∣ Theme]
          What are you writing?; *What's her husband? A film director* (371)
     WHICH [Animacy:personal ∣ nonp, Case=null, Illoc:int, reference:def, role=Referent ∣ Theme]
          Which is your favourite conductor? (Von Karajan or Stokowsky); *Which do you prefer? (Classical or popular music)* (371)
     WHO [Animacy=personal, Case=obj[15] ∣ subj, Illoc=int, reference=indef, role=Agent ∣ Beneficiary ∣ Company ∣ Goal ∣ Referent ∣ Theme]
          Who gave you the present?; *Who has any money?*
     WHOM [Animacy:personal, Case=obj, Illoc:int, reference:indef, role=Beneficiary ∣ Company ∣

---

[14] The numbers in brackets following the examples that illustrate these lexical rules correspond to the page numbers of Quirk et al. (1985) from where these have been retrieved.

[15] Even though Quirk et al. (1985: 370) point out that in objective use *who* is informal and *whom* is formal, they admit that as a prepositional phrase "both *who* and *whom* can take initial position, leaving the preposition at the end of the clause", as in the example *Who(m) is she working for?*

Goal ⎪ Referent]
Who(m) did you give the present to? (818 footnote d)
WHOSE [Animacy:personal, Case=gen, Illoc:int, reference:indef, role=Theme]
Whose is this jacket?

(14) ADVI's Lexical rules:
WHERE [illoc:int, role=Goal ⎪ Location ⎪ Origin ⎪ Position ⎪ Scene]
Where shall I put the glasses?
WHEN [illoc:int, role= Duration ⎪ Frequency ⎪ Time]
When do they meet?
HOW [illoc:int, role=Attribute ⎪ Company ⎪ Condition ⎪ Frequency ⎪ Instrument ⎪ Manner ⎪
Means ⎪ Quantity ⎪ Reason ⎪ Result ⎪ Speed]
How did you mend it?; How did you like her?; How did it go?
WHY [illoc:int, role=Purpose ⎪ Reason]
Why are they always complaining?

(15) DETI's Lexical rules
WHAT [Case=null, Animacy=personal ⎪ nonp, Illoc:int, reference=indef, role=Goal ⎪ Instrument
⎪ Means ⎪ Referent ⎪ Theme]
What conductor do you like best?; *What newspaper do you read?* (369)
WHICH [Case=null, Animacy:personal ⎪ nonp, Illoc:int, reference:def, role=Goal ⎪ Instrument ⎪
Means ⎪ Referent ⎪ Theme]
*Which conductor do you like best? (Von Karajan or Stokowsky); Which newspaper do you
read?(The Times or The Guardian)* (369)
WHOSE [Case=gen, Animacy:personal, Illoc:int, reference:indef, role=Theme]
Whose jacket is this?

In this line, once these categories have been either newly devised or adjusted in order to be able to obtain an accurate encoding of the interrogative PrC-L1-constituent, we consider that the tags ADVI, DETI and PROI (for interrogative adverb, determiner and pronoun, respectively) in ARTEMIS should be introduced in the parsing rules for these fronted WH-forms. Such constituents consist in NPs[16], now referred to as RPs (i.e., our WH-words) that could be regarded as belonging to the two types cited in Cortés-Rodríguez (2016b: 97), "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".[17]

These RPs can be made up of a WH-word as a single nucleus (NUC-RP)[18], that is, a PROI as in the sentence

(16) *Who came yesterday?*

or an ADVI as in the following example:

(17) *Where did you go?*

The nucleus of this fronted component can also be a non-predicative preposition[19], as in

---

[16] Van Valin & LaPolla (1997: 75).
[17] CORE- RP stands for the CORE of the RP; RPIP stands for Initial Position of the RP; and PER-RP stands for Periphery of the RP.
[18] Nucleus of the RP.
[19] When the preposition is predicative, the PP is part of the Periphery (i.e., it is not a core argument, but an optional modifier of the core) and therefore when in initial position it occupies the slot of the Left Detached

To whom did Manuel give the envelope?, where the RP in the PP could again be a PROI or an ADVI. However, both cases can be discarded from the present analysis since they both take part of a LDP[20] and not of a PreC-L1.

Likewise, the interrogative WH-constituent may also be part of an extra core RPIP (Referential Phrase Initial Position) slot and hence regarded as a DETI:

(18) *Which students came to your office?*

This DETI must always be followed by a CORE-RP node with an obligatory NUC-RP (N or Adj) that could be optionally modified by an MP.

(19) *Which books have you lent him?* (818) → DETI RP
(20) *What composers do you like best?* (819) → DETI RP
(21) *Whose beautiful antiques are these?* → DETI MP RP

In turn, this DETI could also take part in a fronted PP like:

(22) *To which students did Manuel give the books?*[21]

An ADVI could also be the constituent in this RPIP, as long as the NUC-RP is an adjective:

(23) *How old are you?*

All this implies having to redesign the parsing rules for phrasal constituents in Cortés-Rodríguez (2016b) as follows:

Modified syntactic rules for phrasal constituents

(24) RP -> RPIP CORE-RP PER-RP || RPIP CORE-RP || CORE-RP PER-RP || CORE-RP || **ADVI** || PRO || PROD || **PROI** || PROP || PROQ || NOUX

(25) RPIP -> PART ART || PART DETP || PART DETD || **ADVI** ||ART || DETP || DETD || **DETI** || RP || MP

(26) NUC-RP-> N || ADJ || ADJ PER-NRP || **ADVI** || PROD || **PROI** || 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

5. PRODUCTION RULES

Three subdivisions inside this section will organize the syntactic rules of our WHQs according to the presence and nature of a 'helping verb' in CORE-initial position. In 5.1 below, the rules elicited illustrate those interrogatives characterized by the presence of a

---

Position (LDP) and not that of the PrCS (see the explanation of the syntactic template selection principle in section 3 above).
[20] See previous note.
[21] Van Valin & Lapolla (1997:175).

single NUC in CORE-initial position. The other two subsections (5.2 and 5.3) will be devoted to the presence of a 'helping verb' (an auxiliary verb or a modal verb) before the NUC in the CORE-node.

The use of auxiliaries represents one of the most resourceful means to form questions in English and they generally divide into primary and modal auxiliaries.[22] Primary auxiliaries, or AUX in this paper, are semantically associated with the grammatical categories of tense, aspect and voice (*be*, *have* and *do*).[23] Modal auxiliaries (*can, could, dare, have to, may, might, must, need, ought to, shall, should, will, would*) are mainly associated with the expression of a modal meaning.

Modal verbs can have a dual nature or modality, and both senses are present in all of them.[24] On the one hand, we have the deontic modality, which could roughly correspond to modes of expressing 'permission', 'obligation' or 'volition'. All these values are encoded in the AVM for the tag MODD proposed for ARTEMIS in Cortés-Rodríguez and Mairal-Usón (2016). Alternatively, we have the epistemic modality, for which Cortés-Rodríguez & Mairal-Usón (2016) propound MODST, a tag in the GDE of our parser, whose AVM encapsulates values like 'possibility', 'necessity' and 'prediction'.

5.1 *WHQs with NUC in CORE-initial position*

The two considerations mentioned above for the formulation of the WHQs syntactic rules within ARTEMIS imply that the most elemental frame for the syntactic rules of WHQs in ARTEMIS has to be a parsing rule with no arguments in the CORE-node (see Fig. 4 below) and with a WH-word which is itself the subject (restricted by [Animacy=personal | nonp, Case=null | subj]) in the extra core slot. That is, a prototypical kernel-1 construction in which the single ARG (the subject) migrates to the PrC-L1 position where it is saturated by a PROI or a DETI, leaving the CORE subjectless and just lodging a single NUC with irrelevant aspect. The NUC of this kernel-1 construction could only be an enclitic negative PRED if *be* is a stative event in a negative WHQ, otherwise an AUX is required (see section 5.2. below).

For ARTEMIS, as opposed to RRG, examples with stative PREDs (*be* or *have*) are regarded as NUCs which may or may not be attended by an ARG[25]. ARTEMIS feeds on FunGramKB's ontological concepts in order to enrich semantically these interrogative frames with primary auxiliaries.[26]
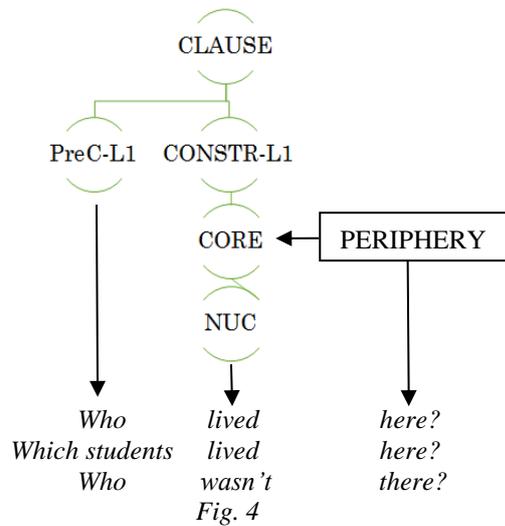
---

[22] Modal verbs are called Modal Auxiliaries by Quirk et al. (1985: 120).

[23] The fact that this type of auxiliaries can also function as main verbs explains the syntactic rules in 5.1 below.
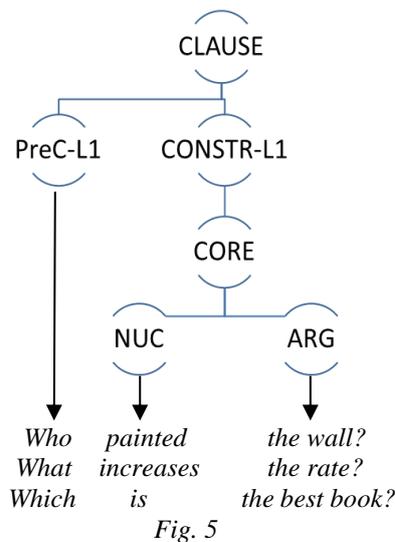
[24] See Quirk et al. (1985: 224ff.) for a further explanation on modality, and Cortés-Rodríguez and Mairal-Usón (2016) and Martín-Díaz (2017) for a translation into RRG terms and a codification within ARTEMIS.

[25] See the example with *be* in Fig. 4 and compare it with that in Fig. 5.

[26] See Martín-Díaz (2017) for a detailed explanation of the dyadic concepts in FunGramKB ([+BE_00], [+BE_01], [+BE_02], [+COMPRISE_00], [+HAVE_00] and [+HOLD_00]) used in the enhanced codification of syntactic rules for YNQs with *be* and *have* as nuclear PREDs in CORE-initial position. These can be also extrapolated to our WHQs.

CLAUSE

PreC-L1   CONSTR-L1

CORE ← PERIPHERY

NUC

Who	lived	here?
Which students	lived	here?
Who	wasn't	there?

Fig. 4

Likewise, in an original kernel-2 construction, one of the semantic arguments can be also transferred to the PreC-L1 position and leaving the CORE subjectless, but occupied by the NUC and the other semantic argument.

CLAUSE

PreC-L1   CONSTR-L1

CORE

NUC   ARG

Who	painted	the wall?
What	increases	the rate?
Which	is	the best book?

Fig. 5

The reduction of core slots in this type of WHQs blocks the existence of kernel-3 constructions as possible syntactic alternatives, since the CORE will always lack a subject that has migrated to the PreC-L1 slot.[27]

---

[27] This rule illustrated in Fig. 6 is similar (except for the PreC-L1 slot) to the one elicited for YNQs in Martín-Díaz (2017) where YNQs with stative nuclear PREDs like *be* and *have* and with two ARGs (*Has she a car?* or *Isn't she a lawyer?*) exemplify the ARG-NUC (or S-O) inversion.

CLAUSE

PreC-L1   CONSTR-L1

CORE

NUC   ARG   ARG

Who   gave   you   the book?
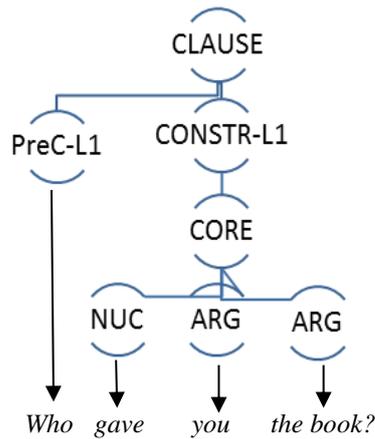
Fig. 6

In other words, the three frames derived out of these interrogatives are always preceded by a PreC-L1 subject (DETI or PROI constrained by the restriction [Animacy=personal | nonp, Case=null | subj]), which can be accompanied by a single NUC with either one or two ARG(s) in the CORE node. Only when an enclitic negative form appears in the input sentence will this NUC be restricted by the use of certain ontological concepts in FunGramKB to a stative PRED.[28]

5.2 *WHQs with AUX in CORE-initial position*

As it happened in our parsing rules for YNQs, the initial position of AUX in the CORE-node of WHQs will indicate the interrogative illocutionary force of the clause, however in this case this AUX is reinforced by the fronted WH-element in the PreCL1 position. The most basic parsing rule for these WHQs is a Kernel-1 construction (Fig. 7 below) in which the subject-argument has been fronted to the extra core slot and the subjectless CORE only lodges an initial AUX (either *be* for progressive aspect or *have* for perfective aspect) followed by NUC.
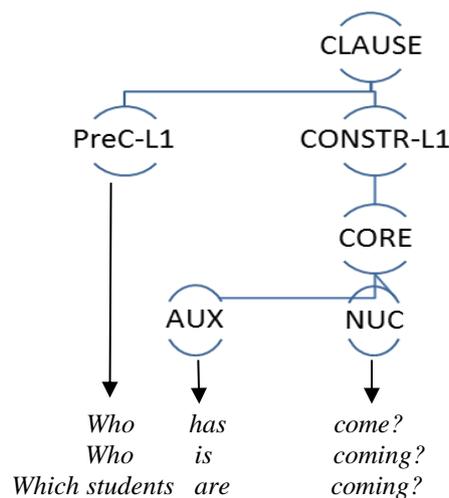

CLAUSE

PreC-L1   CONSTR-L1

CORE

AUX   NUC

Who   has   come?
Who   is   coming?
Which students   are   coming?

Fig. 7

---

[28] See footnote 26.

90

This parsing rule will not only cover the negative structures with the primary auxiliaries *be* and *have* mentioned above, but also all those cases with irrelevant aspect in which *do*-insertion is required (see Fig. 8 below) [29]; that is, including the negative counterparts of those cases in section 5.1. above in which a single NUC in the CORE is used:

Who	*didn't*	*come?*
Who	*hasn't*	*come?*
Who	*isn't*	*coming?*

Fig. 8
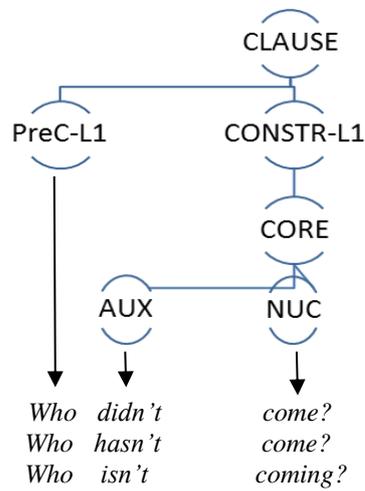
The rest of frames for these WHQs, fronted by other PROI except for the token WHO, inherit the structure of their English YNQ counterparts (Figs. 9-11):

Where	*do*	*you*	*live?*

Fig. 9

---
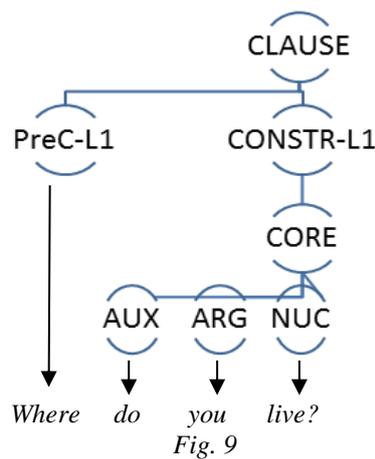
[29] For the relation between *do*-insertion and irrelevant aspect within ARTEMIS see Díaz-Galán and Fumero-Pérez (2016).
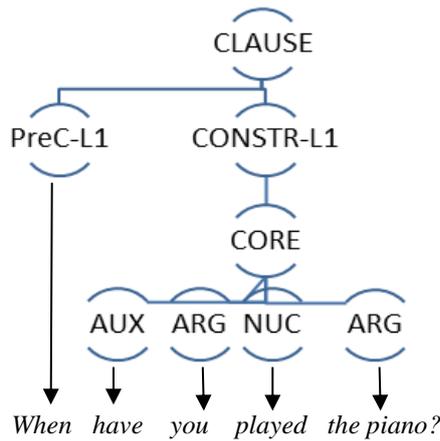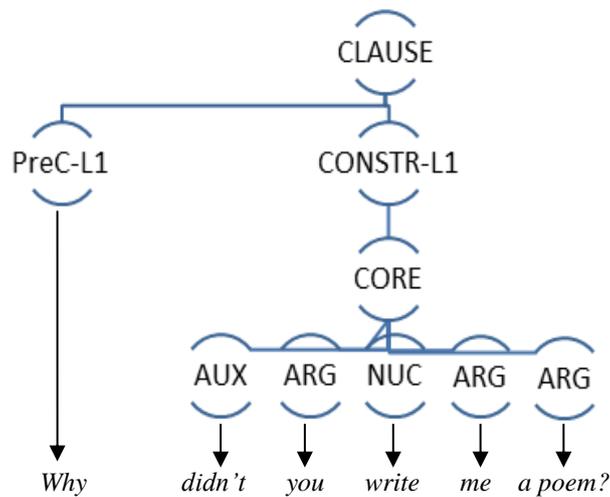
Fig. 10

Fig. 11

5.3 *WHQs with Modals in CORE-initial position*

The Core initial position of our English WHQs can be occupied by either a deontic Modal (MODD), or an epistemic Modal (MODST). The following two subsections 5.3.1. and 5.3.2. will be respectively devoted to their corresponding analysis in ARTEMIS.

5.3.1 *WHQs with MODD in CORE-initial position*
Once again the presence of a PROI or a DETI (restricted by [Animacy=personal ǀ nonp, Case=null ǀ subj]) in the PreC-L1 entails a kernel-1 construction with a subjectless CORE (Fig. 12) which lodges, on this occasion, a single MODD (which can be a positive or an enclitic negative form) and a NUC.

CLAUSE

PreC-L1   CONSTR-L1

CORE

MODD   NUC

Who   *can*   *come?*
Which students *can't*   *come?*
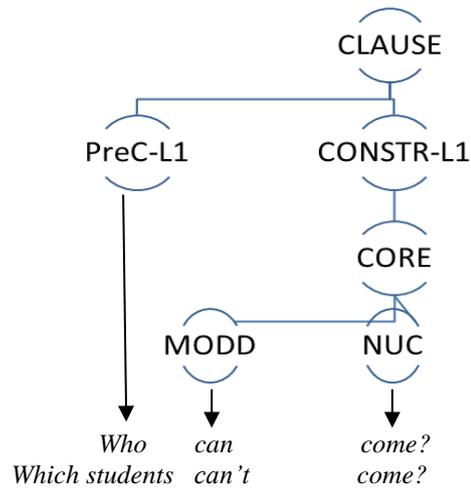
Fig. 12

The rest of frames for this type of WHQs with a fronted WH-component other than WHO will inherit the formulation rules of their English YNQ counterparts with an initial MODD in the CORE node (Figs. 13-15 below):
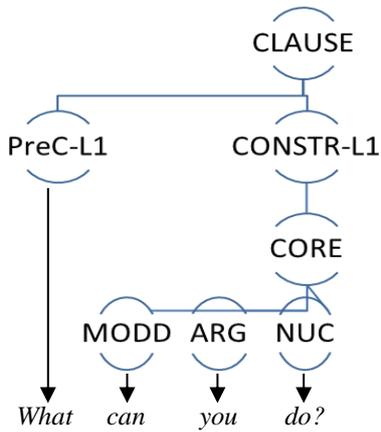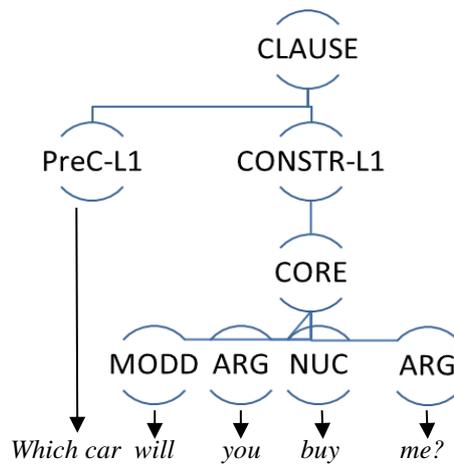

CLAUSE

PreC-L1   CONSTR-L1

CORE

MODD   ARG   NUC

What   *can*   *you*   *do?*

Fig. 13

CLAUSE

PreC-L1   CONSTR-L1

CORE

MODD   ARG   NUC   ARG

Which car *will*   *you*   *buy*   *me?*
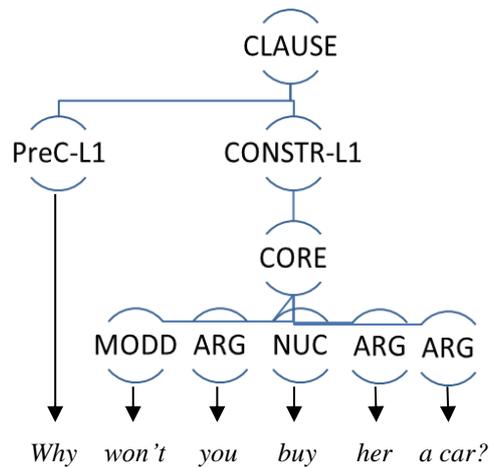
Fig. 14

93

Fig. 15

5.3.2 WHQs with MODST in CORE-initial position

A restricted PROI or DETI in the PreC-L1 also entails here a kernel-1 construction with a subjectless CORE (Fig. 16 below). On this occasion, this CORE lodges a single MODST (which can also be a positive or an enclitic negative form) and a NUC.
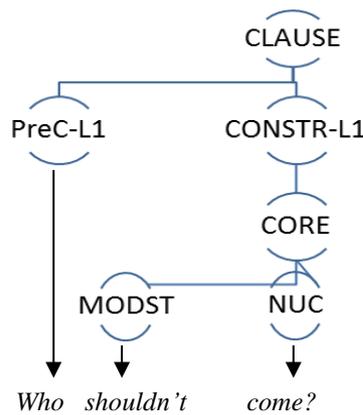

Fig.16

The rest of frames for this type of WHQs with a fronted WH-component other than those restricted by the constraint [Animacy=personal | nonp, Case=null | subj] will inherit the formulation rules of their English YNQ counterparts with an initial MODST in the CORE node (Figs. 17-19):
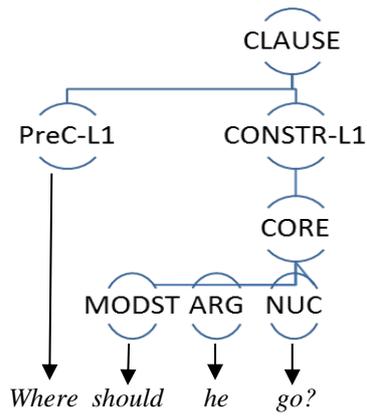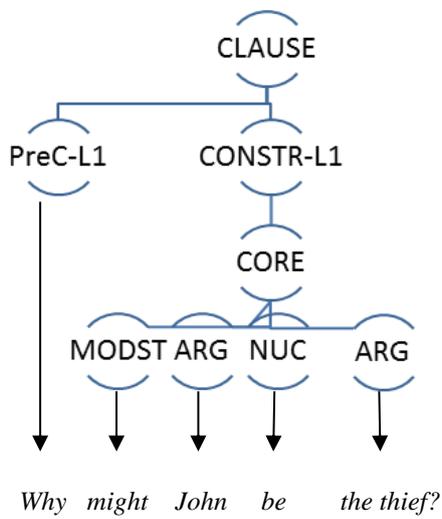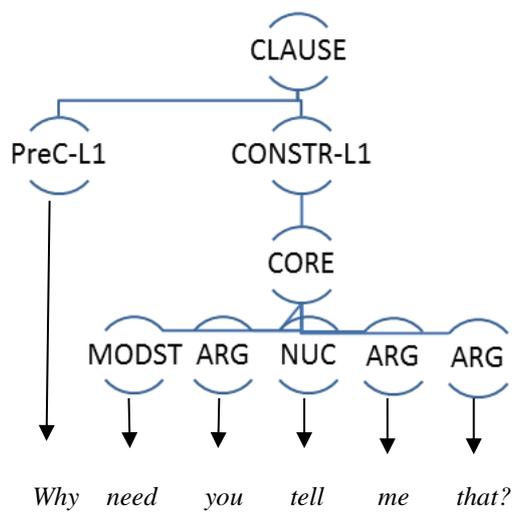
Fig. 17

Fig. 18

Fig.19

6. CONCLUSIONS

The main objective of this paper has been to enable a correct parsing of WH-interrogative structures within ARTEMIS. In order to do that, a series of production rules and AVMs have been proposed inside the GDE of our computer application. Out of these production rules, syntactic rules become particularly relevant since the presence of initial WH-forms in the PCS (or the PreC-L1 position) reveals itself as fundamental for the morphosyntactic analysis of the input sentence ARTEMIS needs to do in order to parse an interrogative. On the other hand, the crucial role that RRG grammatical operators like illocutionary force, modality and status play in the development of an interrogative structure needs to be incorporated to the codification of its constituents. This is done through the selection of a sequence of features that together will conform the AVMs lodging the categories and attributes that will ultimately allow the system to constrain this parsing process (see (8) to (12) in section 4 above).

Likewise, the specific grammatical tokens that introduce these interrogatives (i.e., the different WH-pronouns, WH-determiners, and WH-adverbs) also need to be encoded as instances of the lexical rules in the GDE, where specific values like "case" and "animacy" have also been proposed (see (13) to (15) in section 4 above), contributing thus to the refinement of the necessary constraints.

As a conclusion to the current analysis, the following inventory of CORE-frames is propounded for our WH-interrogatives (see Fig. 20 below). In it, fifteen CORE-frames form the full gamut of syntactic rules for WHQs in English, always preceded by a WH-constituent that can occasionally be restricted to a specific form, as explained above, but where the attribute "illoc" must always be encoded as a feature in the PreC-L1 node that comes to reinforce the interrogative core.

This analysis has enabled us to classify these fifteen CORE-frames as follows: Three of them are headed by a NUC-node where aspect is irrelevant. Out of these three, only one is an argumentless syntactic rule (i.e., a single-NUC CORE-frame). A set of four CORE-frames follows having in common the presence of an AUX in initial position, and where *be* and *have* are aspectual markers and *do* is not (or rather marks an irrelevant aspect). Subsequently, two more sets of four CORE-frames each have been elicited to convey modality in our WHQs, one of these sets is always fronted by a MODD in the linearization of constituents, and the other one is characteristically headed by a MODST in the syntactic sequence.
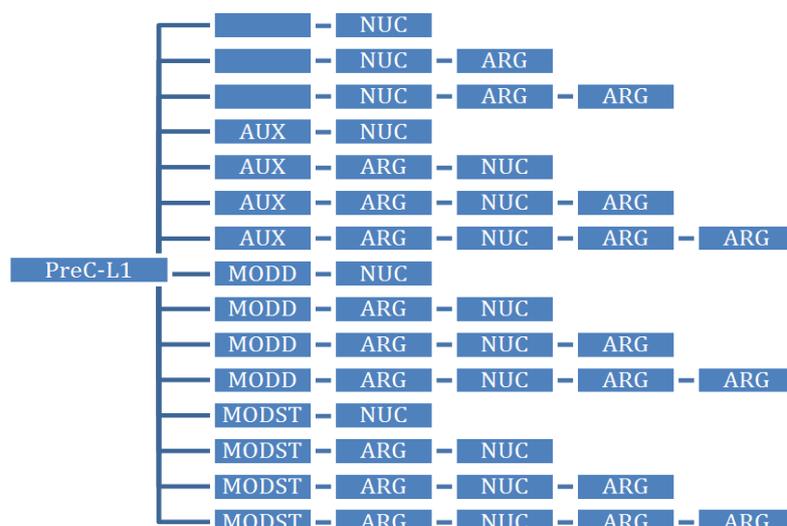

*Fig. 20: **Inventory of CORE-frames for WHQs***

96

We can claim that the catalogue of ten CORE-frames for English YNQs proposed for the GDE in Martín-Díaz (2017) can, almost identically, be proposed in this account for our WHQs.[30] However, five new CORE-frames always headed by a PreC-L1 position have been specifically identified in this paper: NUC; NUC-ARG; AUX-NUC; MODD-NUC; MODST-NUC. All of them imply the absence of a subject argument in the CORE-node of a kernel-1 construction (see examples in Figs. 4, 5, 7, 12 and 16 in section 5 above). This CORE-argument, following the syntactic template selection principle in RRG (Van Valin & LaPolla 1997: 174), is said to migrate to the extra core slot of our syntactic rule, where it plays its interrogative role as a WH-constituent.

In this sense, some new categories (with their corresponding attributes and values) and lexical tokens have been found indispensable in order to show the restrictions that seem to constrain these fronted WH-forms. In this sense, we can conclude that the PreC-L1 position in the above-mentioned five CORE-frames can only be saturated by either a single interrogative pronoun (more specifically, a pronoun constrained by the restriction [PROI: WHAT[31] | WHO[32]]) or a phrasal constituent fronted by an interrogative determiner with the following constraint [DETI: WHAT[33] | WHICH[34]]) and followed by a CORE-RP node.

REFERENCES

Cortés-Rodríguez, F. (2016a). *Parsing simple clauses within ARTEMIS: The computational treatment of the layered structure of the clause in Role and Reference Grammar.* Paper presented at the 34th International Conference of the Asociación Española de Lingüística Aplicada (AESLA). Alicante. April, 14-16.

Cortés-Rodríguez, F. (2016b). 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.

Cortés-Rodríguez, F. & Mairal-Usón, R. (2016). Building an RRG computational grammar. *Onomázein,* 34, 86-117.

Díaz-Galán, A. & Fumero-Pérez, M. C. (2016). Developing parsing rules within ARTEMIS: the case of DO auxiliary insertion. In C. Periñán-Pascual & E. Mestre i Mestre (Eds.), *Understanding Meaning and Knowledge Representation: From Theoretical and Cognitive Linguistics to Natural Language Processing* (pp.283-302). Cambridge: Cambridge Scholars Press.

Mairal-Usón, R. & Ruiz de Mendoza Ibáñez, F. (2008). Levels of description and explanation in meaning construction. In Ch. Butler & J. Martín Arista (Eds.), *Deconstructing Constructions* (pp. 153–198). Amsterdam/Philadelphia: John Benjamins.

---

[30] See footnote 27.
[31] *What increased the rock continent of North America?*
[32] *Who came yesterday?*
[33] *What orchestra conductor is directing?*
[34] *Which students came?*

Martín-Díaz, M. A. (2017). An account of English YES/NO interrogative sentences within ARTEMIS. *Revista de Lenguas para Fines Específicos*, 23 (2), 41-62.

Periñán-Pascual, C. (2013a). Towards a model of constructional meaning for natural language understanding. In B. Nolan & E. Diedrichsen (Eds.), *Linking Constructions into Functional Linguistics: The role of constructions in grammar* [Studies in Language Companion Series 145] (pp. 205–230). Amsterdam: John Benjamins.

Periñán-Pascual, C. (2013b). A knowledge-engineering approach to the cognitive categorization of lexical meaning. *VIAL: Vigo International Journal of Applied Linguistics* 10, 85-104.

Periñán-Pascual, C. & Arcas-Túnez, F. (2010). The architecture of FungramKB. In N. Calzolari, Kh. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner & D. Tapias (Eds.), *Proceedings of the 7th International Conference on Language Resources and Evaluation* (pp. 2667-2674). Malta: European Language Resources Association (ELRA).

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

Periñán-Pascual, C. & Mairal Usón, R. (2011). The COHERENT methodology in FunGramKB. *Onomázein,* 24, 13-33.

Quirk, R., Greenbaum S., Leech, G. & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language.* London: Longman.

Ruiz de Mendoza Ibáñez, F. & Mairal-Usón, R. (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., Wasow, T. & Bender, E. M. (2003). *Syntactic Theory: Formal Introduction.* Stanford: Center for the Study of Language and Information (CSLI) Publications.

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

Van Valin Jr., R. D. & LaPolla, R. (1997). *Syntax: Structure, Meaning and Function.* Cambridge: Cambridge University Press.