



An Account of English YES/NO Interrogative Sentences within ARTEMIS

María Auxiliadora Martín Díaz¹

Universidad de La Laguna

ABSTRACT

As stated in Periñán-Pascual and Arcas Túnez (2010, p. 2671) FunGramKB's lexical and grammatical levels are grounded in sound linguistic theories, allowing the system to capture syntactic-semantic generalizations which are able to provide both explanations and predictions of language phenomena. The linguistic foundation of FunGramKB is inspired on Role and Reference Grammar (RRG, Van Valin, 2005) and the Lexical Constructional Model (LCM, Ruíz de Mendoza & Mairal Usón, 2008; Mairal Usón & Ruíz de Mendoza, 2008). The close interrelation between these grammatical theories and FunGramKB is patent in the design of ARTEMIS (Automatically Representing Text Meaning via an Interlingua-Based SystemTM), an NLP system whose objective is the simulation of natural language understanding. In doing so, ARTEMIS becomes a development of the syntax-to-semantics linking algorithm proposed in RRG which involves a parsing process that starts with a natural language sentence, extracts its morphosyntactic features and provides a representation of these in terms of the so-called layered structure of the clause (LSC) in RRG. Because of its centrality, there is a vital need for a full development of ARTEMIS. This paper aims at contributing to this elaboration by proposing the set of rules and operations necessary for the treatment of Yes/No interrogative sentences within such an NLP system.

Keywords: Yes/No Interrogatives, Role and Reference Grammar, ARTEMIS

1. Introduction

As stated in Periñán-Pascual and Arcas Túnez (2010, p. 2671) the lexical and grammatical levels of Functional Grammar Knowledge Base (FunGramKB) are grounded

¹ **Corresponding author** – Facultad de Filología, Campus de Guajara, San Cristobal de La Laguna 38200, Canary Islands, Spain.

Email: mmartin@ull.edu.es



in sound linguistic theories that allow the system to capture syntactic-semantic generalizations which are able to provide both explanations and predictions of language phenomena.

The linguistic foundation of FunGramKB is inspired on Role and Reference Grammar (RRG, Van Valin, 2005) and the Lexical Constructional Model (LCM, Mairal-Usón & Ruíz de Mendoza, 2008; Ruíz de Mendoza & Mairal-Usón, 2008). The close interrelation between these grammatical theories and FunGramKB is patent 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. In doing so, ARTEMIS becomes a development of the syntax-to-semantics linking algorithm proposed in RRG which involves a parsing process that starts with a natural language sentence, extracts its morphosyntactic features and provides a representation of these in terms of the so-called Layered Structure of the Clause in RRG.

Because of its centrality, there is a vital need for a full development of ARTEMIS. This paper aims at contributing to this elaboration by proposing the set of rules and operations necessary for the treatment of Yes/No interrogative sentences within such an NLP system.

In order to do that, we will firstly introduce the basic structure of FunGramKB, where our main objectives will be settled. A summary of the tenets of importance for this analysis within RRG as well as the fundamentals followed by FunGramKB and ARTEMIS will also be offered here. Section 2 will be devoted to considering the description of Yes/No questions propounded by Quirk et al. (1985, pp. 806-838), together with a brief account of the English modals used in such structures. Section 3 will necessarily be devoted to the treatment of Yes/No questions within RRG. In Section 4, we will deal with the implementation of questions within ARTEMIS. Section 5 will account for the representation of the lexical and syntactic rules of English Yes/No interrogative structures within our NLP prototype. The conclusion in Section 6 will finally display the main achievements attained in this paper.

FunGramKB is a lexico-conceptual modular architecture designed as a knowledge engineering tool for an NLP system capable of simulating natural language comprehension. Its central spine grows along two independent yet interrelated modules. On one side of this structural design we have the linguistic module, a knowledge repository integrated by two constituents, the lexical and the grammatical levels, both characterized by being language specific. On the other side, we have the conceptual module, a more abstract and universal knowledge repository which comprises an ontology, a cognicon and an onomasticon (Periñán-Pascual & Mairal-Usón, 2011).

Following a symbolic-like approach, ARTEMIS comes into play as the NLP system which will exploit FunGramKB as its knowledge base. This prototype is, in turn, theoretically grounded on RRG, a functional linguistic model from which it adopts, among other things, the following fundamental aspects:

- The syntax-semantics linking algorithm as an appropriate bidirectional process to further predict language phenomena.
- The so called Logical Structures (LSs) as a system of representation based on the Aktionsart.
- The Layered Structure of the Clause (LSC), a semantically based formalism where the operator projection, as will be shown later on for our interrogative structures, plays an important role.

In RRG, a representation by layer of the different operators is postulated (Van Valin, 2005) showing how they really work at the different levels within this LSC: i.e., the modality operator functions at the level of the core, whereas the status operator only does so at the clausal level. This distinction will definitely have a computational impact on the design of the production rules and AVMs (Attribute-Value Matrixes) of certain modal verbs within ARTEMIS, as will be observed later on in this section and in Section 4, where a detailed analysis of English interrogative structures will be offered.

Although RRG proved useful for text meaning representation (Periñán-Pascual, 2013), it needed to be enriched with a deeper semantically oriented theory which could also allow us to represent constructional meaning and consequently obtain a finer-grained semantic description out of our computational analysis. In this sense the LCM, a cognitive-functional model, halfway between projectionism and constructivism, was incorporated in order to endow FunGramKB with an enhanced semantic capacity, derived out of the fully-fledged constructional schemata propounded there.

As a system of automatic representation of natural language sentences, ARTEMIS involves a parsing process, which starts by processing the units of a sentence and finishes by extracting their Conceptual Logical Structure or CLSs, enriched versions of RRG's LSs, in which variables are not saturated by lexical units, but by ontological concepts that enable the access to all the conceptual information stored in FunGramKB (i.e., the Ontology, the Cognicon, and the Onomasticon) (Periñán-Pascual & Arcas Túnez, 2014). In line with this, the LSC has been recently extended by the incorporation of a new node, the CONSTR-L1 node, a universal category that will contribute to enhance the lexical meaning of the event as well as that of the CLS (Periñán-Pascual, 2013; Periñán-Pascual & Arcas-Túnez, 2014)

Computationally, this conceptual shift from the LS to the CLS facilitates the mapping into a COREL² scheme, a formalized structure whose metalanguage can be more easily and adequately interpreted by an automated reasoner. Apart from a COREL-scheme Builder, ARTEMIS consists of two more components: The Grammar Development Environment (GDE) and the CLS Constructor (Periñán-Pascual, 2013). It is therefore a modular system that makes possible the access to the required information for an effective parsing. The GDE encapsulates a catalogue of feature-bearing structures, usually represented as Attribute-Value Matrixes (AVMs), and a set of production rules necessary to build a feature-based grammar. The production rules can be of three types: syntactic, constructional and lexical. Whereas syntactic rules need predefining in the GDE in order to assign an adequate parse tree (bracketed or not) to the LSC, constructional and lexical rules “are created in runtime in accordance with the tokens from the input stream” (Periñán-Pascual, 2013, p. 223).

There is a unification process that relates these two constituents of the GDE since each node in the parse tree must be complemented by an attribute-value structure, that is, in this grammar the constituents of the LSC are not regarded as atomic units but as complex meaning-bearing devices that encode selectional and semantic information to be used as constraints on the parsing operations.

In this paper we deal with English Yes/No interrogative structures in particular and, apart from providing their corresponding syntactic rules for a correct parsing to be done in ARTEMIS, we also have to encode the vital information conveyed by the RRG operators (i.e., the illocutionary force, modality and/or status operators) in the so called AVMs. The features lodged in these matrixes will constrain, as will be shown in Section 3, the phrase structure rules of our interrogatives in the LSC.

2. Interrogative sentences: auxiliaries, modals and operators

The analysis of the English Yes/No interrogative structures we will expound in this paper is based on a traditional classification of questions proposed by Quirk et al. (1985, pp. 806-825)³. Two of the fundamental aspects highlighted by these authors in the formation of questions are the inversion of the syntactic order (SVO → VSO) and the use of operators or auxiliaries (primary auxiliaries and modals auxiliaries). A syntactic parser like ARTEMIS will necessarily have to control both of these aspects in

² COREL (COnceptual REpresentation Language).

³ For details, see Quirk et al. (1985, pp. 825 ff.).

order to produce an effective parsing.

Auxiliary and modal verbs can be generally described as “helping verbs” that favour the development of interrogative structures. Their dual nature in English could determine a different syntactic and semantic behaviour, a factor which ARTEMIS should take into account:

- Primary auxiliaries (BE, HAVE and DO) can function as either auxiliaries or main verbs and are semantically associated with the grammatical categories of tense, aspect and voice.
- Modal auxiliaries (CAN, COULD, DARE, HAVE TO, MAY, MIGHT, MUST, NEED, OUGHT TO, SHALL, SHOULD, WILL, WOULD)⁴ always have an auxiliary function and are mainly associated with the expression of modal meaning (possibility, obligation, volition, etc.).

As for the primary auxiliary DO, this is regarded as an empty or ‘dummy’ operator required to form interrogative constructions when there is no other operator present to mark aspect, tense, number and person. However, when the parsing of interrogative structures include the primary auxiliary BE, this can either act:

- As a main verb with a copular function: *Is she pretty?*
- As an auxiliary with a dual behaviour: an aspect auxiliary (*Is Ann learning Spanish?*), or a passive auxiliary (*Was Ann awarded a prize?*).

Likewise, a twofold function for the auxiliary HAVE must also be taken into account:

- As an auxiliary in which it marks perfective aspect: *Has she bought a book?*
- As a main verb with: i. Stative meaning (‘possession’) and syntactic variation: with DO-support (*Do you have a lighter?*), or as an operator itself (*Have you a lighter?*); ii. Dynamic meaning (‘receive’, ‘take’, ‘experience’) and normally combined with DO-support (*Does she have coffee with her breakfast?*).

Modal verbs are classified into two types, according to certain meaning constraints, but both senses are present in all of them. On the one hand, we have the deontic type in which some kind of intrinsic human control over the event is involved: ‘permission’, ‘obligation’, ‘volition’. Quirk et al. call this value deontic modality (1985, pp.

⁴ Modal verbs are called Modal Auxiliaries by Quirk et al. (1985, p. 120) and their use represents one of the most resourceful means to form questions in English. Quirk et al. (1985, p. 137) consider Modality as a gradable property of verbs which can be defined in terms of an “auxiliary verb - main verb scale” in which different categories are identified. Within these only the central modals CAN, COULD, MAY, MIGHT, SHALL, SHOULD, WILL, WOULD and MUST; the marginal modals DARE, NEED and OUGHT TO; and the semi-auxiliary HAVE TO have been studied in the present analysis.

224ff.), and it could roughly correspond to the modality operator in RRG. On the other hand, we have the epistemic type, which typically involves human judgement of what is or is not likely to happen: 'possibility', 'necessity', 'prediction'. Following Quirk et al. (1985, pp. 224ff.) this would be the epistemic modality, which in RRG terms could be related with the status operator.

3. RRG's interrogatives

Since FunGramKB and therefore ARTEMIS are based on RRG as a fundamental grammatical model, some matters on how this grammar treats interrogatives in general, and auxiliaries in particular, need considering.

Van Valin (2005) draws a distinction between an auxiliary (AUX) as a syntactic category and an operator as a grammatical category. For that reason RRG distinguishes a two-fold projection of the clause structure: a constituent projection and an operator projection. Relevant features in the formation of interrogatives such as aspect, tense, modality (deontic and epistemic), and illocutionary force (IF) are usually regarded by RRG as operators, which generally modify the LSC through the operator projection, thus having no syntactic projection as auxiliaries.

In English these features are specifically encoded by the following operators: BE, for progressive: *Is Ann writing a paper?*; HAVE, for perfective: *Have you borrowed my pencil?*; DO, when aspect is irrelevant: *Do you go to school on Saturdays?*; and various modals.

However, BE shows a dual treatment in RRG, since in certain circumstances it is regarded as an AUX of the NUC in the constituent projection (see Figures 1 and 2 below, taken from Van Valin, 2015). According to this author (2005, p. 13, note 3) and Van Valin & LaPolla (1997, p. 115), BE is an AUX in the so-called attributive or identificational constructions (adjectival: *Is she beautiful?*; nominal: *Is that building a hotel?*; prepositional: *Are they in the garden?*), as well as in passive sentences (*Was our team beaten?*).

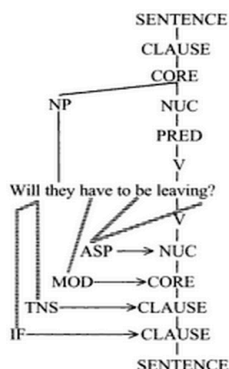


Figure 1. English clause structure.

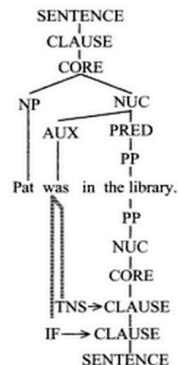


Figure 2. English copular construction.

The notion of operator in Van Valin (2005, p. 13, note 3), however, differs from the same notion in Quirk et al. who define operator as the “first or only auxiliary” which introduces questions, and claim that:

Since BE and sometimes, esp. in BrE, HAVE also have this function as main verbs, the term operator will also be used for them in sentences like: *Is she a tall girl?* or *Has he any money?* (BrE). (1985, p. 79)⁵

This claim may turn convenient when we come to the design of production rules which should be effective enough for ARTEMIS to generate a correct parsing of YES-NO questions. The dual functionality of BE (admitted by Van Valin (2005)), and possibly that of HAVE (evidenced by the previous examples in Quirk et al. (1985)), must have a direct impact on the configuration of AVMs and also on the linearization of the constituents of these syntactic rules, as will be seen in Section 4 below.

4. The implementation of English yes-no questions in ARTEMIS

In the unification process that relates the two constituents of the GDE of our prototype, we will need, as explained in the previous section, the syntactic rules indispensable for the parsing of YNQs⁶ (YES-NO questions) and, on the other hand, the relevant syntactic rules and AVMs that will guarantee a correct parsing of these interrogative structures in English.

Among the basic AVMs we first need to define in our interrogative structures, we have the category of AUX whose tag in the Simple Parts of Speech Tags (POS) of AR-

⁵ Quirk et al. admit that although HAVE can be an AUX in formal British English, HAVE GOT is much more common in this context: *Have you a lighter?* / *have you got a lighter?* (1985, p. 820).

⁶ In the GDE two tags are used for these interrogative structures: SQ for YES/NO Questions and SBARQ for WH-questions. We propose to change them for YNQ and WHQ respectively.

TEMIS was firstly extended by Díaz-Galán & Fumero-Pérez (2016) and further developed by Cortés-Rodríguez & Mairal-Usón (2016). AUX comprises the English auxiliaries BE, DO and HAVE, but in order to include also their enclitic negative forms the attribute for polarity (in bold below) has been incorporated to the original AVM.

```
(1)    <Category Type="AUX">
        <Attribute ID="Aspect"/>
        <Attribute ID="Illoc"/>
        <Attribute ID="Num"/>
        <Attribute ID="Per "/>
        <Attribute ID="Pol"/>
        <Attribute ID="Syn"/>
        <Attribute ID="Tense"/>
    </Category>
```

According to Van Valin, it is the position of the tense operator within the core that triggers illocutionary force in English: “core-medial tense signals declarative IF, core-initial (pre-core) tense signals interrogative IF, and the absence of tense in a matrix core signals imperative IF”⁷. The attribute of tense (<Attribute ID="Tense"/>) in the AVM for the AUX category above seems to be then a relevant feature for our interrogatives, because it is syntactically related to the assignation of the IF of the clause.

Even though Van Valin (2005) seems to be using a notation ambiguous enough not to establish its exact position within the clause, we reckon that it is vital for the development of our syntactic rules within ARTEMIS to do so, and locate with clear precision the category of AUX lodging this tense attribute.

AUX appears as part of the CLAUSE node in Cortés-Rodríguez (2016a and b), immediately preceding the CONSTR-L1 node. However, we have preferred to locate it in CORE initial position. In principle there should be no preference for either option, as there does not seem to exist a case in which the sequence AUX + Subject can be interrupted. This is tantamount to saying that the CORE initial position, the CONSTR-L1 initial position and the position immediately to the left of CONSTR-L1 in the textual sequence are indistinguishable, as shown by the different bracketing possibilities in the following example:

```
(2)    [CLAUSE do [CONSTR-L1 [CORE you often sing in the shower]]]?
```

⁷ “The reason that illocutionary force and tense are linked in the English examples is that, in English illocutionary force is indicated by the position of the tense marker in the main clause: interrogative by core-initial tense, declarative by core-internal tense, and imperative by no tense.” (Van Valin 2005, p. 10, fn. 2; Cf. Van Valin & LaPolla, 1997, p. 42).

[CLAUSE [CONSTR-L1 *do* [CORE *you often sing in the shower*]]]?
 [CLAUSE [CONSTR-L1 [CORE *do you often sing in the shower*]]]?

Our preference for the third bracketing is based on the positional behaviour of BE and HAVE in examples like the following:

(3) *Is she a tall girl?*

*Has he any money?*⁸

Notice that both verbs are analyzed here as tokens of PRED constituents, which typically remain within the CORE layers, even though they are preposed to the left of the subject argument in interrogative structures. Thus, it seems coherent to consider that there is a CORE initial position which can not only be occupied by these types of PREDs, but also by the AUX node when relevant. Note that this seems to be a more economical solution in terms of rule design in our parser, and it complies better with RRG initial description of the positioning of IF operators in the LSC

This category PRED, whose AVM is developed in Cortés-Rodríguez & Mairal-Usón (2016) (see below), becomes consequently relevant for those YNQs in which tokens of the PREDs BE or HAVE are placed in core-initial position in order to mark the interrogative IF. The possibility of having enclitic negative forms in such structures (*Isn't she tall?* or *Hasn't he any money?*) justifies the inclusion of the attribute for polarity in the original AVM designed by the authors above mentioned.

(4) <Category Type="PRED">
 <Attribute ID="Concept"/>
 <Attribute ID="Illoc"/>
 <Attribute ID="Num" />
 <Attribute ID="Per" />
 <**Attribute ID="Pol"**/>
 <Attribute ID="Recip"/>
 <Attribute ID="Reflex" />
 <Attribute ID="Template"/>
 <Attribute ID="Tense"/>
 </Category>

In the implementation of YNQs for ARTEMIS the category of MOD is also relevant. English modal verbs like CAN, COULD, DARE, HAVE TO, MAY, MIGHT, MUST, NEED, OUGHT TO, SHALL, SHOULD, WILL and WOULD should be further specified though in the GDE in relation to the different types traditionally identified as deontic and epistemic modals. As indicated in Cortés-Rodríguez & Mairal-Usón (2016), two tags

⁸ See footnote 4.

and therefore two AVMs, one for the deontic modals (MODD) and the other for the epistemic modals (MODST) are proposed. In both of them is the attribute for polarity included in order to make reference to enclitic negative forms like CANNOT, CAN'T, COULDN'T, WON'T, WOULDN'T, SHOULDN'T, etc.

(5) <Category Type="MODD ">
 <Attribute ID="Illoc"/>
 <Attribute ID="Mod"/>
 <Attribute ID="Num"/>
 <Attribute ID="Per"/>
 <Attribute ID="Pol"/>
 <Attribute ID="Syn"/>
 <Attribute ID="Tense"/>
 </Category>

(6) <Category Type="MODST ">
 <Attribute ID="Illoc"/>
 <Attribute ID="Num"/>
 <Attribute ID="Per"/>
 <Attribute ID="Pol"/>
 <Attribute ID="Sta"/>
 <Attribute ID="Syn"/>
 <Attribute ID="Tense"/>
 </Category>

5. Production rules

The aim of this paper is to give a computational account of English YNQs and, in order to achieve this goal, we need to analyze their syntactic behaviour and design their respective syntactic rules for the GDE in ARTEMIS.

On designing these rules two syntactic nodes in the GDE need consideration, the nuclear and the CORE nodes:

- NUC-node: The presence of a specific PRED (BE or HAVE with a stative meaning) at this level triggers an inversion in the order of the interrogative CORE constituents: NUC(PRED)-ARG-ARG. The lack of aspect at this level triggers in English a specific AUX (DO) in the interrogative CORE.
- CORE-node: An AUX or a MOD (deontic or epistemic) could occupy the CORE initial position in the sequence: AUX ARG NUC or MOD ARG AUX NUC

What happens in the former node seems to have an impact on the latter, therefore

we will concentrate on the development of the production rules for the CORE-node in YNQs.

According to Cortés-Rodríguez (2016a and b), the rule for the CORE node for simple declarative clauses in English (repeated below) subcategorizes three possible frames: Kernel-1, Kernel-2 and Kernel-3 syntactic structures. These frames account for the configuration of the CORE in unmarked cases, as they lodge only the primary predicate of every clause together with its corresponding arguments.

- (7) **CORE**[concept=?, illoc=?, mod=?, neg=?, num=?, per=?, recip=?, reflex=?, tpl=?, t=?] -> **ARG**[concept=?, macro=A | U | n, num=?, per=?, phrase=?, role=agent | attribute | goal | instrument | location | manner | origin | referent | result | theme, tpl=?, var=x | y | w | z] **NUC**[asp=?, concept=?, illoc=?, mod=?, num=?, per=?, recip=?, reflex=?, sta=?, tpl=?, t=?] || **ARG**[concept=?, macro=A | U | n, num=?, per=?, phrase=?, role: agent | attribute | goal | instrument | location | manner | origin | referent | result | theme, tpl=?, var=x | y | w | z] **NUC**[asp=?, concept=?, illoc=?, mod=?, num=?, per=?, recip=?, reflex=?, sta=?, tpl=?, t=?] **ARG**[concept=?, macro=A | U | n, num=?, per=?, phrase=?, role=agent | attribute | goal | instrument | location | manner | origin | referent | result | theme, tpl=?, var=x | y | w | z] || **ARG**[concept=?, macro=A | U | n, num=?, per=?, phrase=?, role=agent | attribute | goal | instrument | location | manner | origin | referent | result | theme, tpl=?, var=x | y | w | z] **NUC**[asp=?, concept=?, illoc=?, mod=?, num=?, per=?, recip=?, reflex=?, sta=?, tpl=?, t=?] **ARG**[concept=?, macro=A | U | n, num=?, per=?, phrase=?, role=agent | attribute | goal | instrument | location | manner | origin | referent | result | theme, tpl=?, var=x | y | w | z] **ARG**[concept=?, macro=A | U | n, num=?, per=?, phrase=?, role=agent | attribute | goal | instrument]

Departing from these intransitive, transitive and ditransitive frames, the following subsections will be devoted to the development of the syntactic rules for YNQs in the GDE. As mentioned in Section 3 above, the dual functionality of the PREDs BE and HAVE will have an impact on the linearization of the constituents of these syntactic rules that must be effective enough for ARTEMIS to generate a correct parsing. This duality to become either the NUC or the AUX of a hypothetical interrogative CORE will be treated in the first two subsections (5.1 and 5.2). On the other hand, Section 5.3 will be devoted to the development of syntactic rules for YNQs whose CORE is headed by modals.

5.1. BE and HAVE as nuclear PREDs in CORE-initial position

A subrule is elicited in the CORE when BE and HAVE (with a stative meaning) are the

verbal predicates (see (8) below). This CORE rule shows inversion, an irrelevant aspect and obviously, interrogative IF, as illustrated in the following sentences.

- (8) NUC ARG ARG
Is she pretty?
Was he in the supermarket?
Has she a car?

Following an approach à la RRG, these structures should be analyzed as comprising an initial AUX be and a subsequent RP/MP/PP which performs as the NUC in a non-verbal predication. This view contrasts with the one proposed for ARTEMIS, that is, a verbal predication in which the stative PRED BE is regarded as a NUC attended by its corresponding ARG(s).

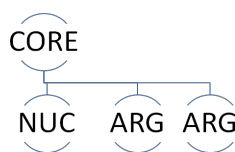
ARTEMIS feeds on FunGramKB's Lexicon and Ontology to enrich these interrogative frames with BE semantically. In particular, three entries of the Lexicon show the different stative events encoded for BE by means of the dyadic concepts: [+BE_00], [+BE_01], [+BE_02].

Likewise, the subrule in (8) also includes interrogatives with the stative event HAVE as a nuclear PRED. ARTEMIS feeds on FunGramKB again in order to retrieve these syntactic rules with HAVE, which is conceptualized in the Ontology through four dyadic entries in the Lexicon: [+HAVE_00], [+BE_01], [+HOLD_00], [+COMPRISE_00].

As a result, the following frame for simple YNQs with the stative events BE or HAVE as nuclear PREDs of the CORE-node are proposed for English, and illustrated with the following graphic examples:

- (9) CORE[concept=?, illoc=int, mod=?, neg=?, num=?, per=?, recip=?, reflex=?, sta=?, tpl=?, t=?] -> **NUC**[asp=null, concept=[+BE_00] [+BE_01] [+BE_02] [+COMPRISE_00] [+HAVE_00] [+HOLD_00], illoc=int, mod=?, num=?, per=?, sta=?, tpl=?, t=?] **ARG**[concept=?, macro=A | U | n, num=?, per=?, phrase=?, role=agent | attribute | goal | instrument | location | manner | origin | referent | result | theme, tpl=?, var=x | y | w | z] **ARG**[concept=?, macro= A | U | n, num=?, per=?, phrase=?, role=agent | attribute | goal | instrument | location | manner | origin | referent | result | theme, tpl=?, var=x | y | w | z]

(10)



*Has she a car?
Isn't she a lawyer?*

5.2. AUX in CORE initial position

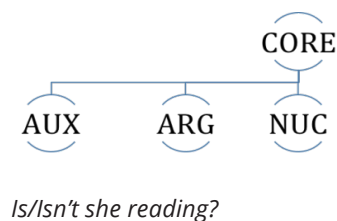
As commented above, it is vital for ARTEMIS to include a node for AUX in the linearization rules within the GDE in order to parse the interrogative CORE accordingly. Our treatment of AUX consequently sides with Unification approaches that distinguish auxiliary from lexical verbs (Sag et al., 2003) and is in line with Díaz-Galán & Fumero-Pérez (2016) who in their analysis for developing DO parsing rules conclude it is an operator that should be located in the CORE-node, even though its scope as an interrogative operator percolates to the clause.

In our parsing rules the initial position of AUX in the CORE-node may imply not only DO-insertion, triggered by the absence of aspect in the NUC (Díaz-Galán & Fumero-Pérez, 2016), but also the inversion of the aspectual operators BE and HAVE.

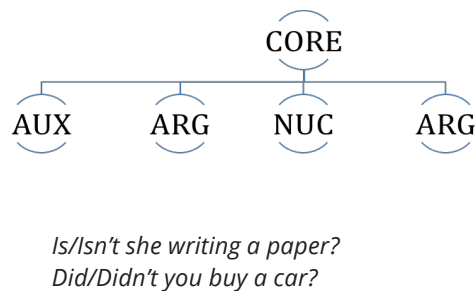
- (11) **CORE** [concept=?, illoc=int, mod=?, neg=?, num=?, per=?, recip=?, reflex=?, sta=?, tpl=?, t=?] -> **AUX**[asp=null | pf | pr, illoc=int, num=pl | sg, per=1 | 2 | 3, t=past | pres] **ARG**[concept=?, macro=A | U | n, num=?, per=?, phrase=?, role:agent | attribute | goal | instrument | location | manner | origin | referent | result | theme, tpl=?, var= x | y | w | z] **NUC**[asp=?, concept=?, illoc=?, mod=?, num=?, per=?, recip=?, reflex=?, sta=?, tpl=?, t=?] || **AUX**[asp=null | pf | pr, illoc=int, num=pl | sg, per=1 | 2 | 3, t=past | pres] **ARG**[concept=?, macro=A | U | n, num=?, per=?, phrase=?, role=agent | attribute | goal | instrument | location | manner | origin | referent | result | theme, tpl=?, var=x | y | w | z] **NUC**[asp=?, concept=?, illoc=?, mod=?, num=?, per=?, recip=?, reflex=?, sta=?, tpl=?, t=?] **ARG**[concept=?, macro=A | U | n, num=?, per=?, phrase=?, role=agent | attribute | goal | instrument | location | manner | origin | referent | result | theme, tpl=?, var= x | y | w | z] || **AUX**[asp=null | pf | pr, illoc=int, num=pl | sg, per=1 | 2 | 3, t=past | pres] **ARG**[concept=?, macro=A | U | n, num=?, per=?, phrase=?, role:agent | attribute | goal | instrument | location | manner | origin | referent | result | theme, tpl=?, var= x | y | w | z] **NUC**[asp=?, concept=?, illoc=?, mod=?, num=?, per=?, recip=?, reflex=?, sta=?, tpl=?, t=?] **ARG**[concept=?, macro=A | U | n, num=?, per=?, phrase=?, role=agent | attribute | goal | instrument | location | manner | origin | referent | result | theme, tpl=?, var=x | y | w | z] **ARG** [concept=?, macro=A | U | n, num=?, per=?, phrase=?, role=agent | attribute | goal | instrument | location | manner | origin | referent | result | theme, tpl=?, var= x | y | w | z]

In them, AUX bears all the morphological features and empties the bracketed description of the PRED of such marks. However, the PRED after each of the aspectual operators BE (AUX [asp=pr]) and HAVE (AUX[asp=pf]) must be either tagged VING[concept=?, tpl=?] for progressive aspect or VPAR[concept=?, tpl=?] for perfective aspect in order to guarantee the necessary presence of a present or past participle form respectively. Besides, the absence of aspect which triggers the insertion of DO (AUX [asp=null]) demands a non-stative event in the PRED of the NUC-node, at least not one related to the concepts [+BE_00] [+BE_01] [+BE_02]⁹.

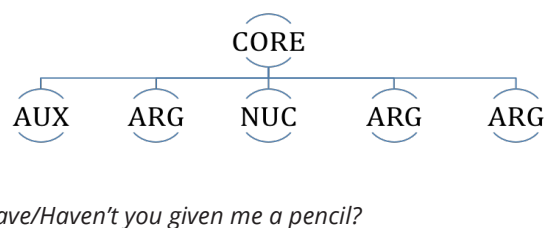
(12)



(13)



(14)



5.3. Modals in the CORE

The syntactic rules for English YNQs including modals basically inherit the bracketed expressions of those with an AUX, but introducing the new constituent with the op-

⁹ The stative PRED HAVE, related to the concepts [+COMPRISE_00], [+HAVE_00] and [+HOLD_00] is commonly combined with DO insertion.

erator MOD.

Since most modals in English show two values (an epistemic meaning of necessity or possibility, and a deontic value of obligation or ability)¹⁰, two different operators have been proposed:

- A core operator for deontic modality (MODD)
- A clausal operator for status or epistemic modality (MODST)¹¹

In this sense, the three-fold development of the original kernel frames is duplicated in order to capture both values.

- (15) **CORE**[concept=?, illoc=int, mod=?, neg=?, num=?, per=?, tpl=?, t=?] ->
MODD[illoc=int, mod=abl | obl | perm | psbl | vol, num=pl | sg, per=1 | 2 | 3, syn=toverb | null, t=past | pres] **ARG** [concept=?, macro=A | U | n, num=?, per=?, phrase=?, role=agent | attribute | goal | instrument | location | manner | origin | referent | result | theme, tpl=?, var=x | y | w | z] **NUC**[asp=?, concept=?, illoc=?, mod=?, num=?, per=?, sta=?, tpl=?, t=?] || **MODST**[illoc=int, num=pl | sg, per=1 | 2 | 3, sta=inf | nec | poss | subj, syn=toverb | null, t=past | pres] **ARG** [concept=?, macro=A | U | n, num=?, per=?, phrase=?, role=agent | attribute | goal | instrument | location | manner | origin | referent | result | theme, tpl=?, var=x | y | w | z] **NUC**[asp=?, concept=?, illoc=?, mod=?, num=?, per=?, sta=?, tpl=?, t=?] || **MODD**[illoc=int, mod=abl | obl | perm | psbl | vol, num=pl | sg, per=1 | 2 | 3, syn=toverb | null, t=past | pres] **ARG**[concept=?, macro=A | U | n, num=?, per=?, phrase=?, role=agent | attribute | goal | instrument | location | manner | origin | referent | result | theme, tpl=?, var=x | y | w | z] **NUC**[asp=?, concept=?, illoc=?, mod=?, num=?, per=?, sta=?, tpl=?, t=?] **ARG**[concept=?, macro=A | U | n, num=?, per=?, phrase=?, role=agent | attribute | goal | instrument | location | manner | origin | referent | result | theme, tpl=?, var=x | y | w | z] || **MODST**[illoc=int, num=pl | sg, per=1 | 2 | 3, sta=inf | nec | poss | subj, syn=toverb | null, t=past | pres] **ARG** [concept=?, macro=A | U | n, num=?, per=?, phrase=?, role=agent | attribute | goal | instrument | location | manner | origin | referent | result | theme, tpl=?, var=x | y | w | z] **NUC**[asp=?, concept=?, illoc=?, mod=?, num=?, per=?, sta=?, tpl=?, t=?] **ARG**[concept=?, macro=A | U | n, num=?, per=?, phrase=?, role=agent | attribute | goal | instrument | location | manner | origin | referent | result | theme, tpl=?, var=x | y | w | z] || **MODD**[illoc=int, mod=abl | obl | perm | psbl | vol, num=pl | sg, per=1 | 2 | 3, syn=toverb | null, t=past | pres] **ARG**[concept=?, macro=A | U |

¹⁰ Van Valin & LaPolla (1997, p. 40, p. 48).

¹¹ Cortés-Rodríguez & Mairal-Usón (2016).

n, num=?, per=?, phrase=?, role= agent | attribute | goal | instrument | location | manner | origin | referent | result | theme, tpl=?, var=x | y | w | z] **NUC**[asp=?, concept=?, illoc=?, mod=?, num=?, per=?, sta=?, tpl=?, t=?] **ARG**[concept=?, macro=A | U | n, num=?, per=?, phrase=?, role=agent | attribute | goal | instrument | location | manner | origin | referent | result | theme, tpl=?, var=x | y | w | z] **ARG**[concept=?, macro=A | U | n, num=?, per=?, phrase=?, role=agent | attribute | goal | instrument] | | **MODST**[illoc=int, num=pl | sg, per=1 | 2 | 3, sta=inf | nec | poss | subj, syn=toverb | null, t=past | pres] **ARG**[concept=?, macro=A | U | n, num=?, per=?, phrase=?, role=agent | attribute | goal | instrument | location | manner | origin | referent | result | theme, tpl=?, var= x | y | w | z] **NUC**[asp=?, concept=?, illoc=?, mod=?, num=?, per=?, sta=?, tpl=?, t=?] **ARG** [concept=?, macro=A | U | n, num=?, per=?, phrase=?, role=agent | attribute | goal | instrument | location | manner | origin | referent | result | theme, tpl=?, var=x | y | w | z] **ARG**[concept=?, macro=A | U | n, num=?, per=?, phrase=?, role=agent | attribute | goal | instrument]

In order to identify the values that modal auxiliaries should codify, Cortés-Rodríguez (2016a and b) has (re)designed the AVMs for the categories of Modality and Status, as well as proposed a series of attributes for their corresponding lexical rules in the GDE.

(16) <Attribute ID="Modality " obl="*" num="1">
 <Value>?mod </Value>
 <Value>ability>abl</Value>
 <Value>obligation>obl</Value>
 <Value> permission>perm</Value>
 <Value> possibility>psbl</Value>
 <Value> volition>vol</Value>
 </Attribute>

(17) <Attribute ID="Status " obl="*" num="1 ">
 <Value>?sta</Value>
 <Value>inference>inf</Value>
 <Value>necessity>nec</Value>
 <Value>possibility>poss</Value>
 <Value>subjunctive>subj</Value>
 </Attribute>

Lexical rules for deontic and epistemic modals:

(18)

MODD

<i>can</i> : [mod:abl, pol:pos, tns:pres]	<i>can't/cannot</i> : [mod:abl, pol:neg, tns:pres]
<i>can</i> : [mod:psbl, pol:pos, tns:pres]	<i>can't/cannot</i> : [mod:psbl, pol:neg, tns:pres]
<i>could</i> : [mod:abl, pol:pos, tns:past]	<i>couldn't</i> : [mod:abl, pol:neg, tns:past]
<i>could</i> : [mod:psbl, pol:pos, tns:past]	<i>couldn't</i> : [mod:psbl, pol:neg, tns:past]
<i>may</i> : [mod:psbl, pol:pos, tns:pres]	<i>might</i> : [mod:psbl, pol:pos, tns:past]
<i>may</i> : [mod:perm, pol:pos]	<i>might</i> : [illoc: int, mod:psbl, pol:pos, tns:past]
<i>must</i> : [mod:obl, pol:pos]	<i>mustn't</i> : [mod:obl, pol:neg]
<i>should</i> : [mod:obl, pol:pos]	<i>shouldn't</i> : [mod:obl, pol:neg]
<i>ought</i> : [mod:obl, pol:pos, syn: toverb]	
<i>oughtn't</i> : [mod:obl, pol:neg, syn: toverb]	
<i>have</i> : [num:sing, per:1 2, mod:obl, pol:pos, syn: toverb]	
<i>have</i> : [num:pl, mod:obl, pol:pos, syn: toverb]	
<i>has</i> : [num:sing, per:3, mod:obl, pol:pos, syn: toverb]	
<i>will</i> : [mod:vol, pol:pos, tns:pres]	<i>won't</i> : [mod:vol, pol:neg, tns:pres]
<i>shall</i> : [mod:vol, pol:pos, tns:pres]	<i>shan't</i> : [mod:vol, pol:neg, tns:pres]
<i>would</i> : [mod:vol, pol:pos, tns:past]	<i>wouldn't</i> : [mod:vol, pol:neg, tns:past]
<i>should</i> : [mod:vol, pol:pos, tns:past]	<i>shouldn't</i> : [mod:vol, pol:neg, tns:past]

MODST

<i>may</i> : [sta:poss, pol:pos, tns:pres]	<i>might</i> : [sta:poss, pol:pos, tns:past]
<i>must</i> : [sta:nec, pol:pos]	<i>mustn't</i> : [sta:nec, pol:neg]
<i>should</i> : [sta:inf, pol:pos]	<i>shouldn't</i> : [sta:inf, pol:neg]
<i>ought</i> : [sta:inf, pol:pos, syn: toverb]	
<i>oughtn't</i> : [sta:inf, pol:neg, syn: toverb]	
<i>have</i> : [num:sing, per:1 2, sta:nec, pol:pos, syn: toverb]	
<i>have</i> : [num:pl, sta:nec, pol:pos, syn: toverb]	
<i>has to</i> : [num:sing, per:3, sta:nec, pol:pos, syn: toverb]	
<i>needn't</i> [sta:nec, pol:neg]	<i>may</i> [sta:subj, pol:neg]

The modals that are currently registered in ARTEMIS are CAN, CANNOT, DARE, MAY, MUST, OUGHT, SHALL, SHOULD and WILL, but, as can be seen, other tokens have been added and encoded by Cortés-Rodríguez (2016a and b) in the GDE. For example, NEED, which has been included because, especially in British English, the rare epistemic value of MUST in questions may be covered by this nonassertive modal auxiliary¹²:

- (19) Need they make all that noise?

Likewise, even though the modal construction of DARE only occurs in non-assertive structures¹³, the values identified by Quirk et al (1985, p. 138) might be also registered in the list encoded by Cortés-Rodríguez (2016a and b) above, because encod-

¹² NEED TO with DO-support is more common, though: *Do they need/have to make all that noise?*

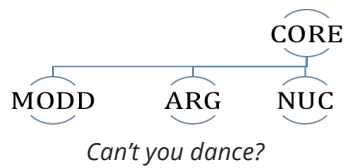
¹³ It only occurs occasionally in non-assertive form in British English, a context in which it is usually replaced by the main verb *dare* and, especially in AmE, by the blend construction with *do* plus the bare infinitive (cf. Quirk et al., 1985, 3.42).

ing attributes like prediction and permission (Quirk et al 1985, p. 138) in our parsing rules for YNQs would serve to explain the following distinction:

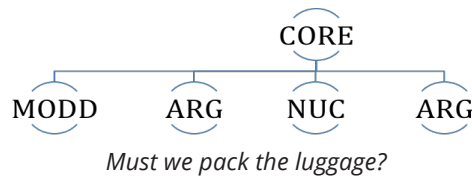
- (20) Prediction → *Dare we escape?* (138)
 Permission → *Dare I suggest a compromise between your two positions?* (816)

Here we have an illustration of the possible frames for Modal insertion in English YNQs

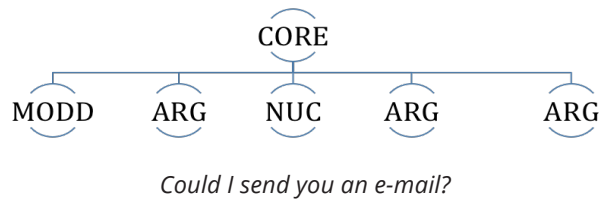
(21)



(22)



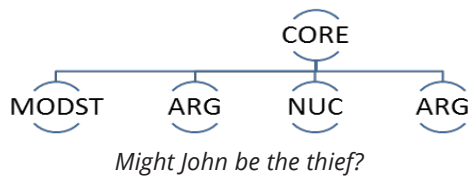
(23)



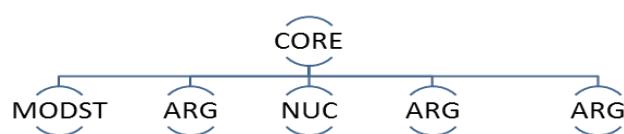
(24)



(25)



(26)

*Need I remind you that?*

6. Conclusions

This paper aims at contributing to the development of ARTEMIS as a vital NLP tool for the simulation of natural language understanding. In general, it has implied the analysis of interrogative structures, and in particular revisiting one of RRG's tenets in the attempt to provide the GDE with a set of appropriate production rules necessary for the correct parsing of YNQs.

The following catalogue of ten interrogative core frames (Figure-3 below) has been postulated for ARTEMIS. They represent the possible syntactic alternatives for YNQs in English. Only one of these frames, the first one in this catalogue, illustrates the ARG – NUC inversion with stative nuclear PREDs like BE and HAVE and with two ARGs. AUX insertion characterizes the following three core-frames (where AUX are DO for neutral aspect, BE for progressive aspect and HAVE for perfective aspect). Finally, six of these parsing rules show MOD insertion, that is, three frames for each type of modals, three for MODDs and three for MODSTs.

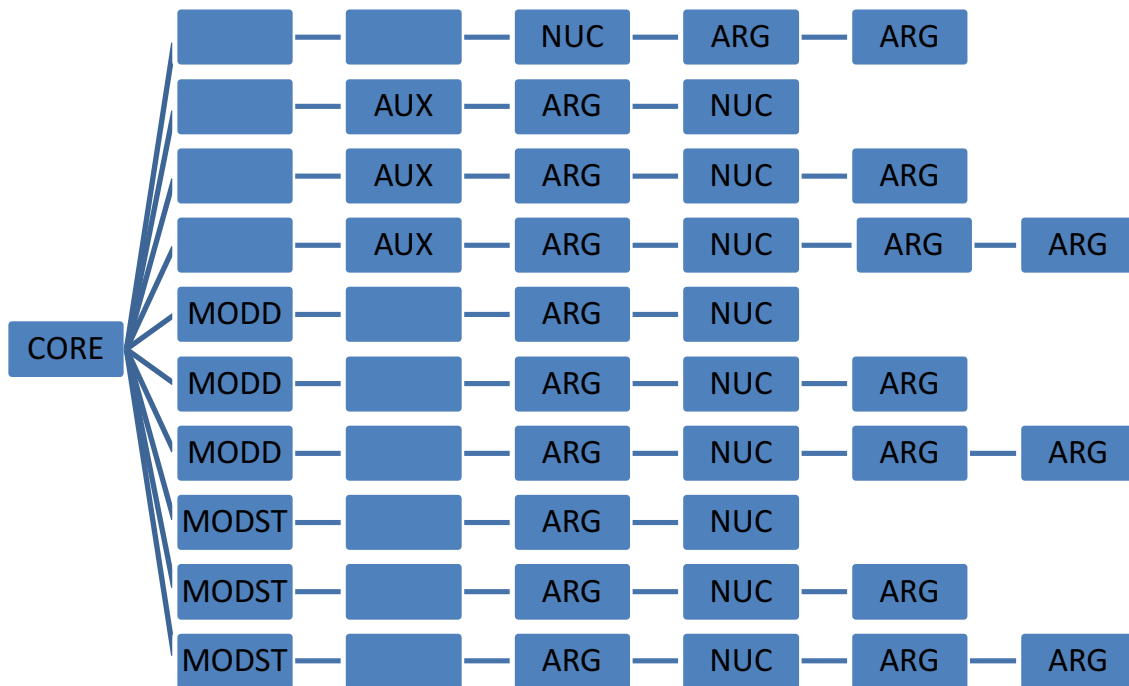


Figure 3. Catalogue of CORE frames for YNQs.

Article history

Paper received: 23 February 2017

Paper received in revised form and accepted for publication: 23 May 2017

About the author

María Auxiliadora Martín Díaz is an Associate Professor of English Language and Linguistics in the Department of English and German Philology, University of La Laguna, Spain. Her areas of research interest are the interaction between lexis and grammar within functional and cognitive models, and diachronic lexicology. She has been a research member in several research projects dealing with these areas, and has supervised several PhD dissertations on these topics. She is currently involved in the project entitled

Desarrollo de un laboratorio virtual para el procesamiento computacional de la lengua desde un paradigma funcional (FF2014-53788-C3-1-P) funded by the Spanish Ministry of Science. She is also a member of the research group GILGLI (*Research Group on English Lexis and Grammar*) that forms part of the NeuroCog Research Network, co-funded by the *Agencia Canaria de Investigación, Innovación y Sociedad de la Información* and ERDF.

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. *34th International Conference of AESLA*. Alicante, 14-16, April 2016.
- 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. Perrián-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.
- Perrián-Pascual, C. (2013). 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
- Perrián-Pascual, C. & Arcas-Túnez, F. (2010). The architecture of FungramKB. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*, (pp. 2667-2674). Malta: European Language Resources Association (ELRA).
- Perrián-Pascual, C. & Arcas-Túnez, F. (2014). The implementation of the FunGramKB CLS constructor, In B. Nolan & C. Perrián-Pascual (Eds.), *Language Processing and Grammars* (pp. 165-196). Amsterdam: John Benjamins.
- Perrián-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, R. D., Jr. (2005). *Exploring the syntax-semantics interface*. Cambridge: Cambridge University Press.
- Van Valin R.D. Jr. & LaPolla, R. (1997). *Syntax: Structure, meaning and function*. Cambridge: Cambridge University Press.