# ARTEMIS: parsing non-peripheral complex sentences in ASD-STE100[1]

## Marta González-Orta
Universidad de La Laguna
España

## Auxiliadora Martín-Díaz
Universidad de La Laguna
España

**Marta González-Orta:** Departamento Filología Inglesa y Alemana, Universidad de La Laguna, España.
| E-mail: maorta@ull.edu.es
**Auxiliadora Martín-Díaz:** Departamento Filología Inglesa y Alemana, Universidad de La Laguna, España.
| E-mail: mmartind@ull.edu.es

ONOMÁZEIN 56 (June 2022): 80 - 99
Marta González-Orta and Auxiliadora Martín-Díaz
ARTEMIS: parsing non-peripheral complex sentences in ASD-STE100

81

# Abstract

The simulation of natural language understanding is one of the main objectives of natural language processing (NLP). Within the different applications designed for this purpose, the ARTEMIS prototype follows the paradigm of unification grammars, and unlike other trending computational resources, it is theoretically grounded in Role and Reference Grammar (RRG). The syntax-to-semantics linking algorithm proposed in this functional grammar lies at the basis of a parsing process that starts with a natural language sentence, extracts its morpho-syntactic features and provides a representation of these in terms of the so-called layered structure of the clause (LSC).

The Grammar Development Environment (GDE) in ARTEMIS is a major component where feature-based production rules (syntactic, lexical and constructional) are stored and ready to allow the generation of the enhanced layered structure of the clause of natural language expressions. Syntactic rules that account for phrasal constituents and simple sentences have already been described, but it is now turn to focus on the study of non-peripheral complex sentences. In an attempt to validate these syntactic rules and to avoid some of the common problems that may arise in parsing applications, our research will concentrate on the analysis of RRG's juncture-nexus combinations as found in a controlled natural language (CNL) such as ASD-STE100 (Aerospace and Defence Simplified Technical English).

**Keywords:** non-peripheral sentences; ARTEMIS; production rules; ASD-STE100.

ONOMÁZEIN 56 (June 2022): 80 - 99
Marta González-Orta and Auxiliadora Martín-Díaz
ARTEMIS: parsing non-peripheral complex sentences in ASD-STE100

82

# 1. Introduction

The structure of this paper derives out of one of the main objectives of our research project, the validation of the production rules for the Grammar Development Environment (or GDE module) in ARTEMIS (Automatically Representing Text Meaning via an Interlingua-based System). In order to do that, we have opted for studying non-peripheral complex structures as found in a controlled language, since this will necessarily provide a more constrained grammar to work with and therefore simplify the mechanisms to finally evaluate our prototype.

Consequently, in section 1 ASD-STE100, the simplified controlled language to be used in our research, will be introduced. Secondly, a summary of the most important aspects of complex sentences within Role and Reference Grammar (RRG—Van Valin and LaPolla, 1997; Van Valin, 2005) will be considered, since ARTEMIS is inspired on this functionally-oriented linguistic theory. Thirdly, this paper will show the influence of the Lexical Constructional Model (LCM—Mairal-Usón and Ruiz de Mendoza, 2008) on FUNK Lab, a virtual laboratory for natural language processing, where the computational resources FunGramKB (Functional Grammar Knowledge Base) and ARTEMIS, used in this research, are encountered (Mairal-Usón and Cortés-Rodríguez, 2017). Section 4 is devoted to the analysis *à la* ARTEMIS of the non-peripheral complex sentences identified in ASD-STE100. Finally, with the intention of supplying the GDE with the necessary tools to carry out a correct parsing of our complex sentences, a series of attribute-value matrixes (AVMs) and production rules (lexical, syntactic and constructional) will be proposed to be validated in a forthcoming research. The concluding remarks of this research are presented in section 5.

# 2. Controlled Languages and ASD-STE100

ASD-STE100 stands for Aerospace and Defence Simplified Technical English, and is often abbreviated to STE or just Simplified English. It is a controlled natural language (CNL—Kuhn, 2014) developed for the readability of maintenance documentation of the aerospace and defence industries of Europe, to make their texts more uncomplicated and less condensed than when full English is used. It had its origins in 1979, even though it did not receive its current name until 2005 when AECMA (*Asociación Española de Contruction Management*) merged with two other associations to form ASD. According to the authors of their website (http://www.asd-ste100.org/), the success of STE is such that even industries not related to this discipline use it beyond its original purpose, thus stimulating a growing interest in academic, scientific and professional circles on the linguistic side.

The ASD-STE100 guide (January 2017 version, or issue 7) is based on English, but its restrictive general rules (to be developed in section 4 below) constrain the language at different levels: lexical, syntactic and semantic. Within these it is obviously the syntactic level which

ONOMÁZEIN 56 (June 2022): 80 - 99
**Marta González-Orta and Auxiliadora Martín-Díaz**
ARTEMIS: parsing non-peripheral complex sentences in ASD-STE100

83

will greatly restrict our research on complex sentences, especially if the intention behind this is automatic translation.

## 3. RRG's non-peripheral juncture-nexus combinations in English

The four layers identified in RRG, the nuclear, the core, the clausal and the sentential layers (figure 1, taken from Van Valin and LaPolla, 1997: 38), correspond as well with the four levels of juncture that characterize complex structures in this grammatical model.

**FIGURE 1**

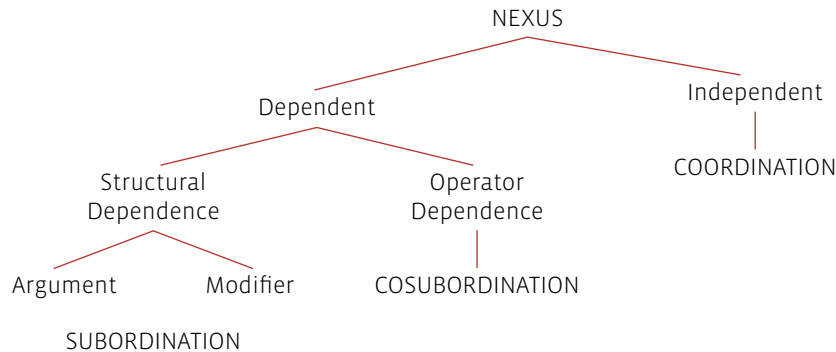Abstract LSC including extra-core slots and detached positions[2]
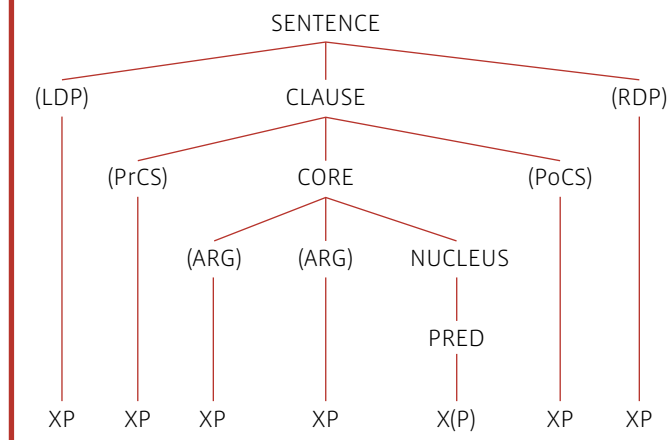


In these four layers of junctures three types of nexus relations can be found: coordination, cosubordination and subordination. The first type, coordination, concerns independent structures, whereas the other two, cosubordination and subordination, involve a structure or an operator dependency (figure 2 from Van Valin and LaPolla (1997: 454)). In Van Valin and LaPolla (1997: 469), only coordination was admitted in the sentential level, since cosubordination had no sentential operators to share, and subordination had no sentential units to be embedded. Van Valin (2005: 192-193), however, claims that sentential subordination can also be possible. This proposal has already been illustrated for ASD-STE100 for temporal and concessive structures (in Martín-Díaz, 2019)[3].

---

2    Van Valin (2005) incorporates a higher dominating TEXT-node when describing sentential complex structures (see sections 3.8 and 5.7 below).

3    Sentential subordination examples in ASD: *Before a new autobrake mode engages, the active selection disengages* or *Although the pressure decreases, the valve must stay close.*

ONOMÁZEIN 56 (June 2022): 80 - 99
**Marta González-Orta and Auxiliadora Martín-Díaz**
ARTEMIS: parsing non-peripheral complex sentences in ASD-STE100

84

**FIGURE 2**
Nexus types



As a result of these considerations, eleven are the possible juncture-nexus combinations attested in the languages of the world according to RRG. Out of these, nuclear coordination and nuclear subordination have been discarded in English (Van Valin and LaPolla, 1997: 454-455). Besides, in a study of non-peripheral complex sentences like the present paper, we also need to dispose of the sentential subordination juncture-nexus type because of its obvious adverbial nature (Martín-Díaz, 2019). As a conclusion, only the eight combinations below will be the scope of this research:

(1) Nuclear cosubordination
   Core coordination
   Core cosubordination
   Core subordination
   Clausal coordination
   Clausal cosubordination
   Clausal subordination
   Sentential coordination

In RRG the presence of CLMs (clause-linkage markers) contributes to the categorization of juncture-nexus combinations. Following Van Valin and LaPolla (1997: 476), "languages have a category of what we will call clause-linkage markers which serve to express important aspects of the syntax and semantics of complex constructions". They may include conjunctions and switch-reference markers. Accordingly, *to* for example is a core-level CLM which links cores. However, when the dependent unit is a clause, as in an English object complement, *that* is the clause-level CLM (Van Valin (2005: 205))[4]. Finally, CLMs will also be present at the

---

4    Many adverbial subordinate clauses are also introduced by CLMs like the English subordinators *because, if* or *although* (see Martín-Díaz, 2019), but they will not be the object of study in this paper.

ONOMÁZEIN 56 (June 2022): 80 - 99
Marta González-Orta and Auxiliadora Martín-Díaz
ARTEMIS: parsing non-peripheral complex sentences in ASD-STE100

85

levels of the sentence and text, where 'text' is described as "the highest node dominating two sentence nodes" (Van Valin, 2005: 192), as section 4 below will show.

## 3.1. Nuclear cosubordination in English

"Nuclear junctures are single cores containing more than one nucleus […] taking a single set of core arguments" (Van Valin and LaPolla, 1997: 448). Out of the three possible nexus combinations at this nuclear level, only cosubordination has been identified for English. The relevant operators to be shared in it are nuclear directionals, nuclear negation and nuclear aspect. This type of juncture does not permit a complementizer and the second nucleus must be intransitive, that is, an intransitive verb, adjective or preposition taking a single argument, because the use of a transitive verb would create a CORE juncture (Van Valin and LaPolla, 1997: 446). The two nuclei in these structures permit a variable ordering: they may be adjacent, as in *Push open the door*[5], or separated, as in *Push the door open*. This duality can only occur when the second nucleus contains a state predicate which may be adjectival or prepositional, as in *He pushed the table over* or *He pushed over the table*. However, even in these circumstances can the alternative be constrained to the weight of the NP (see (2) below): "the heavier the NP, the more acceptable the examples with adjacent nuclei become" (Van Valin and LaPolla, 1997: 446).

(2) *Bill pushed the door closed.*
    *\*Bill pushed closed the door.*
    *Bill pushed closed the heavy door that had just been repainted after the storm.*

## 3.2. Core coordination in English

According to Van Valin and LaPolla (1997: 448), in core coordination a single CLAUSE contains more than one CORE. However, argument sharing characterizes this type of juncture since one argument is syntactically and semantically present in one of the CORES, but only has a semantic function in the linked CORE. As already mentioned above, a grammatical marker or complementizer, the so-called CLM, is usually required to indicate this linked unit within the clause (Van Valin and LaPolla, 1997: 470; Van Valin, 2005: 205).

In general, CORE coordination is employed for jussive[6], direct perception and propositional attitude (Van Valin and LaPolla, 1997: 481) and we can illustrate it with the following examples:

---

5    All examples in this section have been taken from Van Valin and LaPolla (1997), except when specified otherwise.

6    "The expression of a command, request or demand (Lyons 1977), e.g. *Pat asked the students to leave, the king ordered the troops to attack the city*" (Van Valin and LaPolla, 1997: 479).

ONOMÁZEIN 56 (June 2022): 80 - 99
Marta González-Orta and Auxiliadora Martín-Díaz
ARTEMIS: parsing non-peripheral complex sentences in ASD-STE100

86

(3) *Dana saw Chris washing the car.*
   *John must try to wash the car.*

In CORE coordination, the CORE modality operator *must* in *John must try to wash the car* only has scope over the first CORE of the LSC (i.e., *John must try*) and not over the second (*to wash the car*). ARTEMIS, as a linear syntactic parser, needs this operator to become a constituent in the production rules of the GDE. As such, this operator has been labelled MODD or MODST depending on the type of modality concerned and defined in terms of the feature-based structures that ARTEMIS uses to constrain its parsing, the so-called AVMs (see section 3 for a development)[7].

## 3.3. Core cosubordination in English

CORE cosubordination instantiates aspectual[8], psych-action[9] and purposive[10] relations (Van Valin and LaPolla, 1997: 481). Argument sharing also characterizes core cosubordination. Nevertheless, in this nexus type CORE nodes are dominated by a superordinate CORE node (Van Valin, 2005: 203), where the CORE-level operators are shared across all the nuclei (i.e., the deontic *must* in the following example):

(4) *Carlos must wash the car and clean his room.*

Van Valin and LaPolla (1997: 460) argue that despite the presence of the conjunction *and*, there is no coordination in this CORE juncture since the scope of the modal *must* is not only over the first CORE but also over the second one with which it shares the semantic macrorole actor (i.e., *Carlos*).

## 3.4. Core subordination in English

Structural dependence characterizes core subordination in RRG. Two nexus types of juncture can take place: daughter and peripheral subordination. The former applies to CORE arguments realized by subject complement clauses, canonically constrained to gerunds and

---

7   In ARTEMIS, the label MODD for deontic modals has been assigned to differentiate these from epistemic modals, labelled MODST (Cortés-Rodríguez, 2016).

8   "A separate verb describes a facet of the temporal envelope of a state of affairs, specially its onset, its termination or its continuation, e.g. *Chris started crying, Fred kept singing, Harry finished writing the chapter*" (Van Valin and LaPolla, 1997: 479).

9   "A mental disposition regarding a possible action on the part of a participant in the state of affairs, e.g. *Max decided to leave, Sally forgot to open the window, Tanisha wants to go to the movies*" (Van Valin and LaPolla, 1997: 479).

10  "One action is done with the intent of realizing another state of affairs, e.g. *Juan went to the store to buy milk, Susan brought the book to read*" (Van Valin and LaPolla, 1997: 479).

ONOMÁZEIN 56 (June 2022): 80 - 99
**Marta González-Orta and Auxiliadora Martín-Díaz**
ARTEMIS: parsing non-peripheral complex sentences in ASD-STE100

87

subject *that*-complement clauses. The latter applies to modifier phrases (relative and adverbial clauses), as illustrated below.

(5) *Washing the car today would be a mistake.*
*That she arrived late shocked everyone.*
*I liked the cars which were destroyed yesterday.*
*John saw Max after he went to the party.*

Both subtypes, daughter and peripheral subordination, happen to be discarded from our analysis of ASD non-peripherals. On the one hand, because it is obvious that a study of non-peripheral structures must exclude modifier phrases; on the other hand, because of the constraints imposed on the guidelines of ASD-STE100. The restrictions in this controlled language specify: i. that you can use "the '-ing' form of a verb only as a modifier in a technical name" (p. 1-3-4); ii. that you can only use the conjunction THAT "after verbs such as 'make sure', 'show' and 'recommend'" (p. 1-9-9). The second of these restrictions implies that only the object *that*-complement subtype is allowed in ASD, a juncture-nexus combination that concerns the CLAUSE layer, and not the CORE (see sections 3.7 and 5.6 below).

## 3.5. Clausal coordination in English

The defining feature of this universal juncture-nexus combination is the impossibility of sharing an operator, which involves that each clause may have a distinct illocutionary force. This is illustrated in the example below, where the first clause is an imperative and the second an assertion, both connected by the conjunction *and* (Van Valin, 2005: 199):

(6) *Sit down and I'll fix you a drink.*

## 3.6. Clausal cosubordination in English

Clausal cosubordinate juncts exhibit clausal operator dependence, that is, tense and/or illocutionary force must be shared across all juncts and therefore governed by a superordinate CLAUSE-node as shown in figure 11 below. This juncture-nexus combination satisfies the requirements of the conjunction reduction construction (Van Valin and LaPolla, 1997: 521-522; Van Valin, 2005: 230), characterized by "a sequence of events sharing a common primary topical participant" (Van Valin and LaPolla, 1997: 522).

(7) *Paul drove to the store and bought some beer.*
*Robin drove out of Phoenix this morning and will arrive in Atlanta tomorrow.*

## 3.7. Clausal subordination in English

Subordinate juncts at the level of the clause have no argument sharing and operator dependence is not significant for them (Van Valin and LaPolla, 1997: 457). They function either as

ONOMÁZEIN 56 (June 2022): 80 - 99
Marta González-Orta and Auxiliadora Martín-Díaz
ARTEMIS: parsing non-peripheral complex sentences in ASD-STE100

88

daughter or as peripheral subordinate clauses. The second subtype of clausal subordination concerns adverbial or peripheral clauses as treated in Van Valin (2005: 194). The daughter subordination subtype, on the other hand, expresses propositional attitude[11], cognition[12] and indirect[13] relations. According to Van Valin (2005: 199-200), this juncture implies a "syntax-semantics mismatch" that "violates the basic principle that arguments in the logical structure of the verb are realized as core arguments". Examples in (8) below illustrate this subtype in which the embedded clauses are semantically regarded as an argument of the matrix verb, but syntactically considered to occur outside the core.

(8) *Frank said that his friends were corrupt.*
    *Paul considers Carl to be a fool.*
    *Kim told Pat after work that she will arrive at the party late*[14].

## 3.8. Sentential coordination in English

Two complete sentences with their corresponding left-detached positions (LDP), *As for Sam* and *as for Paul*, as seen in (9), conform this linkage (Van Valin, 2005: 192). This position "is outside of the clause but within the sentence" (Van Valin and LaPolla, 1997: 36). Consequently, in sentential coordination, a TEXT-node will dominate two subordinated SENTENCE-nodes, as shown in figure 13 below.

(9) *As for Sam, Mary saw him last week, and as for Paul, I saw him yesterday.*

## 4. ARTEMIS and the parsing of non-peripherals

Within FunGramKB NLP tools, ARTEMIS is a prototype application among others[15] designed with the aim of enabling the understanding of natural languages in the framework of RRG and constraint-based grammars. As a knowledge base, FunGramKB allows ARTEMIS to do an

---

11   "The expression of a participant's attitude, judgement or opinion regarding a state of affairs" (Van Valin and LaPolla, 1997: 479).

12   "An expression of knowledge or mental activity" (Van Valin and LaPolla, 1997: 479).

13   "An expression of reported speech" (Van Valin and LaPolla, 1997: 479).

14   Van Valin (2005: 199) illustrates with this example the fact that "English does not allow phrasal peripheral elements to occur between two core elements, and consequently because the peripheral PP 'after work' occurs between 'Pat' and the 'that-clause', the embedded clause must be outside of the core".

15   The Laboratory of Natural Language Processing and Text Analytics (NLP-LAB); the FunGramKB NAVIGATOR; a resource for discovering and extracting terminology (DEXTER); and the application Data Mining Encountered (DAMIEN).

ONOMÁZEIN 56 (June 2022): 80 - 99
**Marta González-Orta and Auxiliadora Martín-Díaz**
ARTEMIS: parsing non-peripheral complex sentences in ASD-STE100

89

effective parsing by providing a knowledge repository where rigorous morphosyntactic, semantic and pragmatic information is offered.
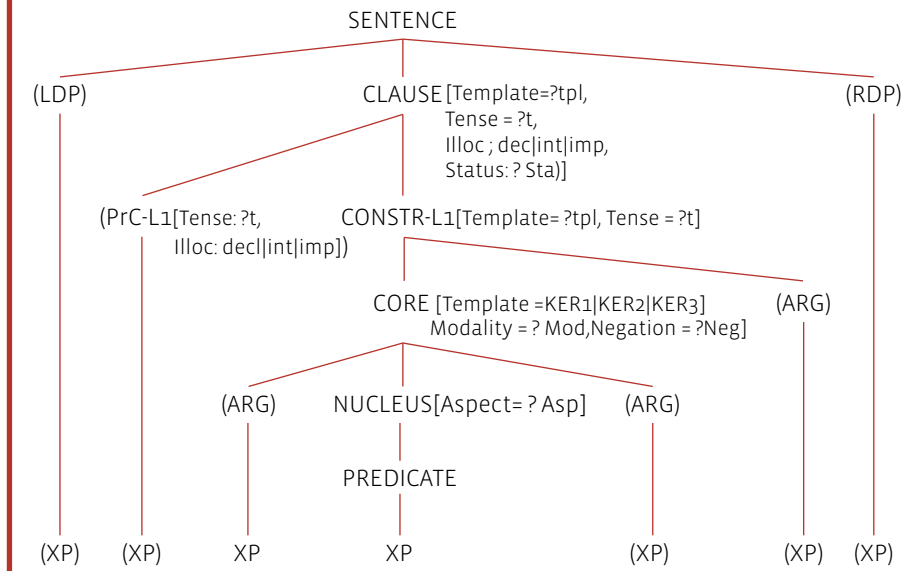
ARTEMIS consists of three modules that enable the encoding of natural-language sentences: the CLS constructor, a tool for the generation of a conceptual logical structure; the COREL-scheme Builder, a tool to transform a CLS into a connceptual representation language and make ARTEMIS useful for NLP tasks; and the GDE, the Grammar Development Environment, where the grammar building process takes place. Three types of production rules, lexical, syntactic and constructional, conform the GDE along with a catalogue of constraining attribute-value matrixes (AVMs), i.e., "complex formal descriptions of grammatical units" that contribute to the parsing process (Periñán- Pascual, 2013).

Production rules in ARTEMIS are intended to computationally enrich the framework of RRG's LSC and its linking algorithm. Syntactic rules derive from the Lexicon in FunGramKB, where kernel structures are described and stored:

(10)  intransitive or kernel-1: *It rained.*

(11)  monotransitive or kernel-2: *Mary ate a sandwich.*

(12)  ditransitive or kernel-3: *Mary gave John the book.*

Construction rules include the non-kernel constructions that derive from FunGramKB's Grammaticon. In this module, four different levels (argumental, or L1; implicational, or L2; illocutionary, or L3; and discursive, or L4) mirror the multilevel constructional view of the Lexical Constructional Model (LCM). These non-kernel constructions can be even recursively generated with the aid of the verbs' core grammar together with all its constructional schemata (for example, the level 1 transitive-resultative construction *John kicked the ball flat*, or the level 1 caused motion construction *John kicked the ball into the stadium*) (Periñán-Pascual, 2013: 214).

The above mentioned kernel/non-kernel distinction involved the introduction of a new constituent in the LSC of RRG, the CONSTR-L1 node, between the CORE and the CLAUSE nodes, and, in turn, the redefinition of RRG's precore slot position as a preconstruction-L1-position. The LSC is thus reinterpreted as one or more L1-constructions where "the innermost construction introduces the core, which can be modeled by other L1-constructions, typically contributing with a further argument" (Periñán-Pascual, 2013: 222). The introduction of this constructional L1 node has two implications for the analysis of complex structures in ARTEMIS. On the one hand, it will allow us to reinterpret RRG's nuclear cosubordination, since the CONSTR-L1 node comes to supply the secondary NUC (NUC-S) that will modify the CORE (figures 6 and 7). On the other, RRG's core junctures characterizing complex structures should be redefined as L1 junctures in ARTEMIS. Both adjustments are shown in figure 3 below.

ONOMÁZEIN 56 (June 2022): 80 - 99
**Marta González-Orta and Auxiliadora Martín-Díaz**
ARTEMIS: parsing non-peripheral complex sentences in ASD-STE100

90

**FIGURE 3**

Modified LSC in ARTEMIS[16]



Taking into account that ARTEMIS needs to follow a linear processing in its syntactic and constructional rules, each of the constituents in the linear sequence will be assigned a structural slot in the LSC. Unlike RRG's tree diagrams, this consideration implies the proposal of a CLM as a direct node in the constituent projection, as proposed in the parsing of adverbial clauses in Martín-Díaz (2019). In fact, RRG's CLMs have been recently relabeled as LMs (linkage markers) in ARTEMIS because the new CONSTR-L1 layer added to the original three-tier proposal in Van Valin's LSC demands LMs at said layer too.

Three different realizational possibilities have been propounded in the lexical rule (13) below (Cortés-Rodríguez, 2019) where LMs could be lexicalized by complementizers (COMP), coordinators (COORD) or subordinating conjunctions (CONJ):

(13)   LM → COMP || COORD || CONJ

LMs introducing non-peripheral complex structures are lexicalized by a COMP or functional words such as *to* or *that* or a COORD or logical connectors such as *and*, *or*, *but* that allows ARTEMIS to trigger a possible complex template from the Grammaticon and semantically identify it with a specific type of L4-construction (see figure 4 below). CONJs, otherwise, introduce CONSTR-L1 peripheral constituents (González-Orta and Martín-Díaz, 2019; Martín-Díaz, 2019).

---

16   The ARG node derived from the mothernode CONSTR-L1 constitutes the NUC-S of the nuclear cosubordination subtype.

**FIGURE 4**
L4-Constructicon in FunGramKB Editor



The subsequent AVM derived from the consideration of this LM category is based on Martín-Díaz (2018), where two attributes "Nexus" and "Syn" were proposed. Whereas the attribute for "Nexus" applies to our analysis, the attribute "Syn", developed in turn by Cortés-Rodríguez (2016), is irrelevant since there are no cases of syntactic co-occurrence in our data gathering[17]. Instead Cortés-Rodríguez (2019) suggests the attribute "Juncture" to account for the type of dependent unit that is conjoined, developed accordingly in (14-16) below:

(14) ‹Category type= "LM"›
      ‹Attribute ID="Nexus"/›
      ‹Attribute ID="Juncture"/›
      ‹/Category›

(15) ‹Attribute ID="Nexus" obl="+" num="1"›
      ‹Value›?nexus‹/Value›
      ‹ValueTag="subordinating"›sub‹/Value›
      ‹ValueTag="cosubordinating"›cosb‹/Value›
      ‹ValueTag="coordinating"›coord‹/Value›
      ‹/Attribute›

(16) ‹Attribute ID="Juncture" obl="+" num="1"›
      ‹Value›?junc‹/Value›

---

17 By syntactic co-occurrence we mean sentences in which the presence of a given LM implies a specific syntactic form, that is, cases like *have to* + infinitive or *be* + v-*ing* or *after/before* + v-*ing*.

ONOMÁZEIN 56 (June 2022): 80 - 99
Marta González-Orta and Auxiliadora Martín-Díaz
ARTEMIS: parsing non-peripheral complex sentences in ASD-STE100

92

‹ValueTag="clause"›cl‹/Value›
‹ValueTag="CONSTR-L1"› CONSTR-L1‹/Value›
‹ValueTag="sentence"›s‹/Value›
‹/Attribute›

## 5. Parsing non-peripherals in ASD-STE100

Seven non-peripheral juncture-nexus combinations have been identified in ASD-STE100. Consequently, not all the juncture-nexus combinations explained for English in section 2 above are present in our controlled language. Mostly, our seven juncts are characterized by the participation of an LM-node. Only the nuclear level is distinguished by its absence. As we can infer from our corpus, such a constituent must be present in the three other layers: CONSTR L-1, CLAUSE and SENTENCE, as the tree diagrams in the subsequent sections illustrate. The lexical realizations of the LM in these non-peripheral juncts in ASD are

(17)  CONSTR-L1-level: (*not*) *to* [junc=CONSTR-L1; nexus=coord|cosb]
      CLAUSE-level: *that* [junc=cl; nexus=sub]
      CLAUSE-level: *and, or* [junc=cl; nexus=coord|cosb]
      SENTENCE-level: *and, or, but* [junc=s; nexus=coord]

## 5.1. Nuclear cosubordination in ASD-STE100

This juncture-nexus combination can be illustrated in our controlled language with causative and resultative constructions like the ones exemplified below. As such, these constructions are considered L1-constructions and therefore stored in FunGramKB's L1-constructicon (see figure 5 below).

**FIGURE 5**
L1-Constructicon in FunGramKB Editor

ONOMÁZEIN 56 (June 2022): 80 - 99
**Marta González-Orta and Auxiliadora Martín-Díaz**
ARTEMIS: parsing non-peripheral complex sentences in ASD-STE100

93

The tree-representations for each of them, along with their corresponding syntactic rules, are shown below (figures 6 and 7):

**FIGURE 6**
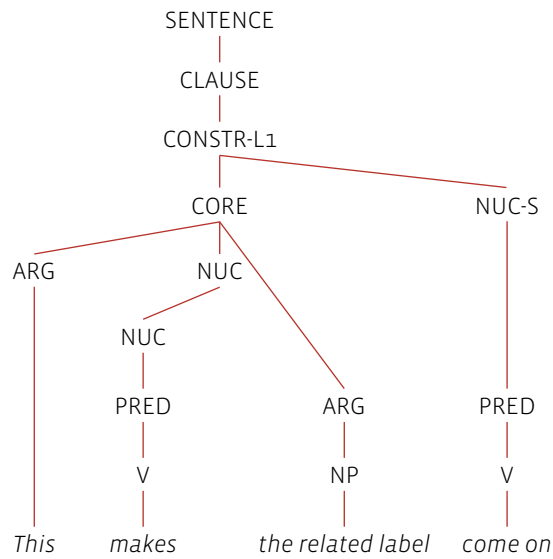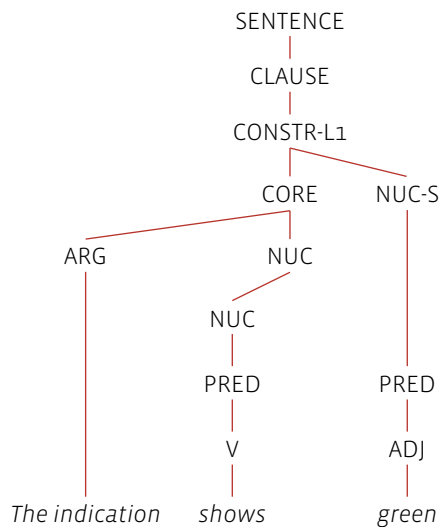Constituent projection of NUCLEAR cosubordination (Verbal NUC-S)



**FIGURE 7**
Constituent projection of NUCLEAR cosubordination (Adjectival NUC-S)



(18)  Nuclear cosubordination [LM: Ø]
      CONSTR-L1 → CORE-NUC-S,
      where NUC-S → V|| ADJ

ONOMÁZEIN 56 (June 2022): 80 - 99
**Marta González-Orta and Auxiliadora Martín-Díaz**
ARTEMIS: parsing non-peripheral complex sentences in ASD-STE100

94

## 5.2. CONSTR-L1 coordination in ASD-STE100

CONSTR-L1 coordination in our controlled language is employed for commands, requests or demands. As can be seen in the tree diagram below, these juncts are generally characterized by having two CONSTR-L1s coordinated by a CONSTR-L1-level LM, which is lexicalized, as seen in the syntactic rule (19) by (*not*) *to*.

**FIGURE 8**

Constituent projection of CONSTR-L1 coordination
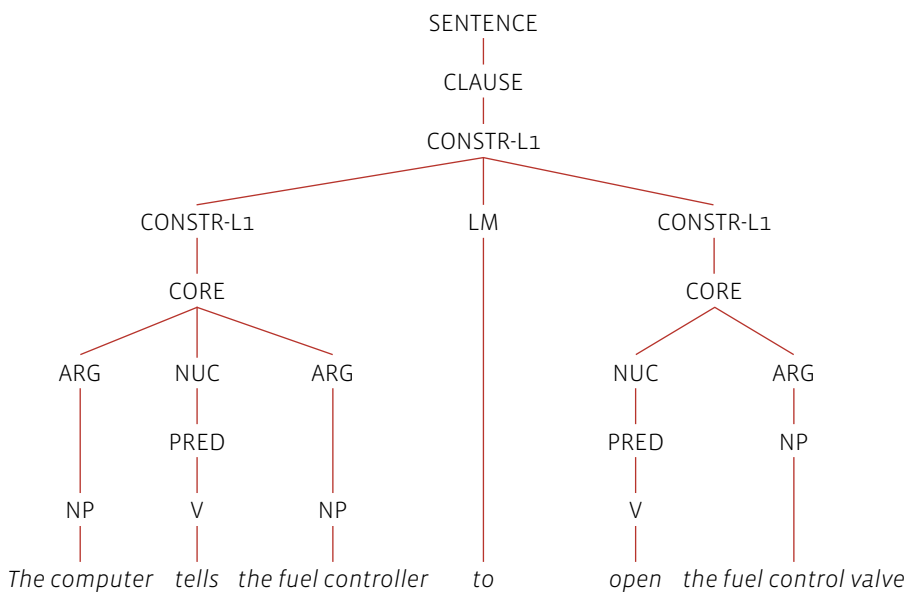


(19)  CONSTR-L1 coordination [LM: to | not to]
       CONSTR-L1 → CONSTR-L1-LM- CONSTR-L1,
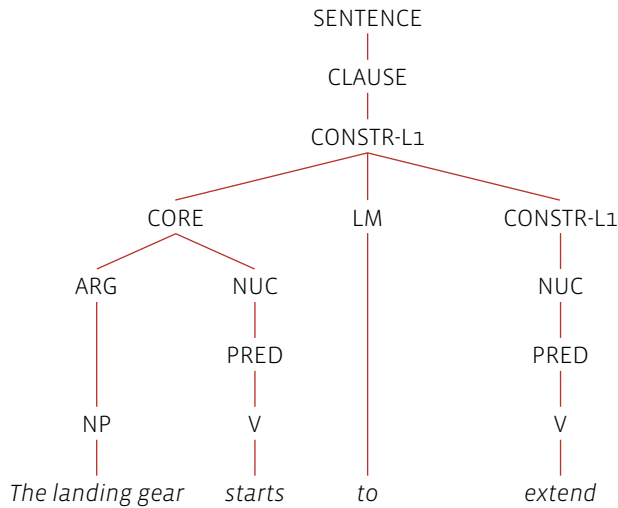       where LM → COMP

## 5.3. CONSTR-L1 cosubordination in ASD-STE100

CONSTR-L1 cosubordination in ASD-STE100 is instantiated by aspectual and psych-action relations[18]. In this type of CONSTR-L1 cosubordination, CONSTR-L1 nodes are thus dominated by a superordinate CONSTR-L1 node and cosubordinated by a CONSTR-L1-level LM *to*.

---

18    As an example of aspectual relation, see figure 9. As an example of psych-action relation: [*it*] *tries to do the pre-land test again with the other side*.

ONOMÁZEIN 56 (June 2022): 80 - 99
Marta González-Orta and Auxiliadora Martín-Díaz
ARTEMIS: parsing non-peripheral complex sentences in ASD-STE100

95

**FIGURE 9**

Constituent projection of CONSTR-L1 cosubordination
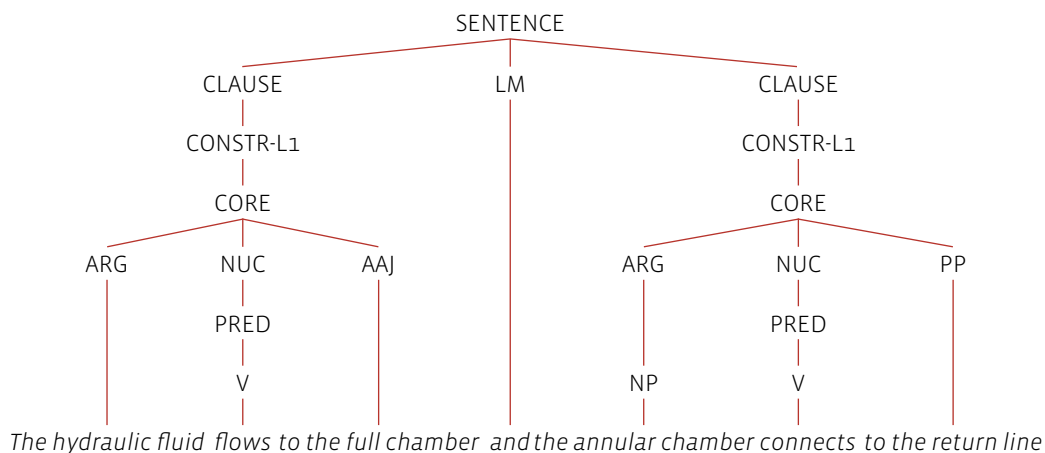


(20)  CONSTR-L1 cosubordination [LM: to]
      CONSTR-L1 → CORE-LM- CONSTR-L1
      where LM → COMP

## 5.4. Clausal coordination in ASD-STE100

In ASD-STE100, the three types of clausal juncture-nexus combinations (coordination, cosubordination and subordination) are possible. In particular, CLAUSAL coordination has a SENTENCE-level LM that coordinates two CLAUSE-nodes. As we can see in rule (21), such an LM is a COORD lexicalized by *and*, *but*, *or*.

**FIGURE 10**

Constituent projection of CLAUSAL coordination

ONOMÁZEIN 56 (June 2022): 80 - 99
Marta González-Orta and Auxiliadora Martín-Díaz
ARTEMIS: parsing non-peripheral complex sentences in ASD-STE100

96
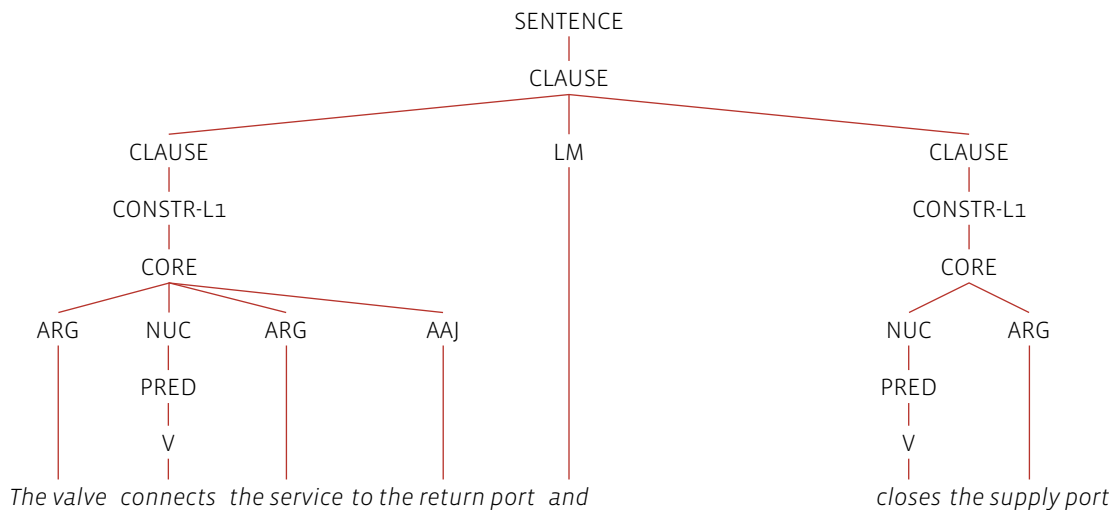
(21)  Clausal coordination [LM: and | but | or]
     S → CL-LM-CL,
     where LM → COORD

## 5.5. Clausal cosubordination in ASD-STE100

A superordinate CLAUSE-node governs the CLAUSE-level LM that cosubordinates the two CLAUSE nodes in figure 11. Only examples of conjunction reduction are available for this juncture-nexus combination in our CNL (see section 2.6). The CLAUSE-level LM lexicalizes the conjunctions *and, or,* as seen in the syntactic rule (22).

**FIGURE 11**
Constituent projection of CLAUSAL cosubordination



(22)  Clausal cosubordination [LM: and | or]
     S → CL,
     where CL → CL-LM-CL,
     where LM → COORD

## 5.6. Clausal subordination in ASD-STE100

Out of the two subtypes that characterize this juncture-nexus combination, the only possible subtype (the object *that*-complement daughter subordination) expresses knowledge or mental activity in ASD-STE100[19]. As seen in figure 12, a matrix CLAUSE-

---

19   The other subtype of clausal subordination, that of peripheral or adverbial clauses, has already been treated in Martín-Díaz (2019).

ONOMÁZEIN 56 (June 2022): 80 - 99
Marta González-Orta and Auxiliadora Martín-Díaz
ARTEMIS: parsing non-peripheral complex sentences in ASD-STE100

97

node lodges a CLAUSE-level LM COMP *that* which, in turn, introduces a subordinated CLAUSE functioning as argument of the matrix unit. The lexical rule for the ARG-node reads as follows and its realizational possibility can be twofold: the first will account for subordinated clauses functioning as objects, and the second for those as subjects (-*ing* clauses not found in STE):

(23)   ARG → LM-CLAUSE || CLAUSE

**FIGURE 12**
Constituent projection of CLAUSAL subordination
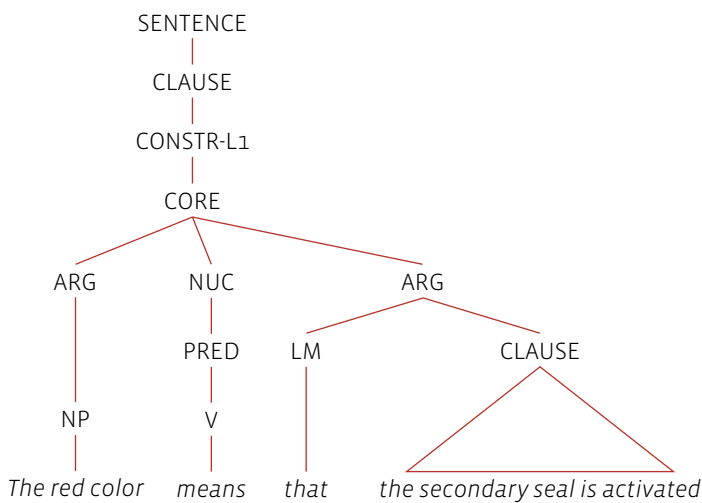


(24)   Clausal subordination [LM: that]
         CONSTR-L1 → CORE,
         where CORE → ARG-NUC-ARG
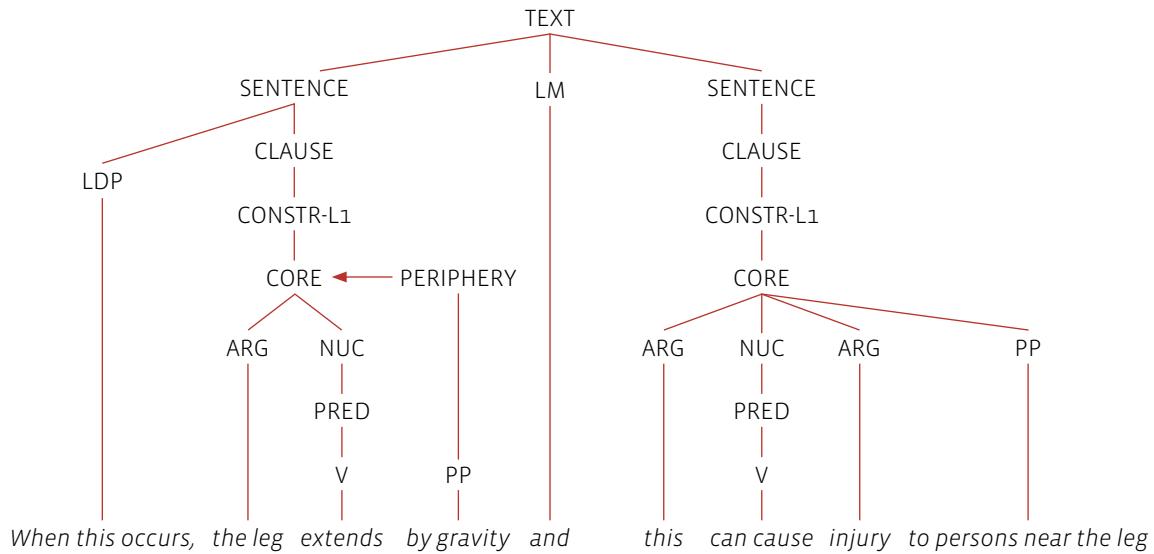         where ARG → LM-CLAUSE
         where LM → COMP

## 5.7. Sentential coordination in ASD-STE100

In the sentence layer, only sentential coordination is available in ASD-STE100. In this type of linkage two complete sentences, or SENTENCE-nodes, are linked by a dominating TEXT-node. The first of these sentences includes a LDP as seen in figure 13 and a TEXT-level LM COORD *and*, *or*.

(25)   Sentential coordination [LM: and | or]
         TEXT → S-LM-S,
         where LM → COORD

ONOMÁZEIN 56 (June 2022): 80 - 99
**Marta González-Orta and Auxiliadora Martín-Díaz**
ARTEMIS: parsing non-peripheral complex sentences in ASD-STE100

98

**FIGURE 13**
Constituent projection of SENTENTIAL coordination



## 6. Conclusion

The analysis of non-peripheral complex sentences in our CNL ASD-STE100 has been approached as a way of constraining the study of complex structures in English. This has involved the description of seven juncture-nexus combinations: cosubordination at the nuclear level; coordination and cosubordination at the CONSTR-L1 level; coordination, cosubordination and subordination at the clause level; and, finally, coordination at the sentence level.

Despite the nuclear cosubordination examples found in ASD, most of our junctures require the presence of a linkage marker. This fact has taken us to introduce some modifications in RRG as a consequence of its necessary adaptation to the linearization principles observed in our parsing application. These modifications have resulted in a reinterpretation of RRG's CLM as a direct LM node in ARTEMIS. Consequently, from our data gathering a three-layered LM has been proposed, i.e., CONSTR-L1-level, CLAUSE-level and SENTENCE-level LM. As our syntactic rules for non-peripherals have shown, this LM category will be lexicalized by complementizers or coordinators which will allow the three different nexuses to occur.

The analysis carried out in the present paper has provided two important contributions for the enrichment of the GDE component: 1. a set of syntactic rules for the seven combinations abovementioned; 2. the development of constraining AVMs for the category LM (mainly the attributes "Nexus" and "Juncture"). This will contribute to the parsing process of complex structures in English, enabling the interaction between the syntactic rules for non-peripherals and the L4-constructional templates in the Grammaticon.

ONOMÁZEIN 56 (June 2022): 80 - 99
Marta González-Orta and Auxiliadora Martín-Díaz
ARTEMIS: parsing non-peripheral complex sentences in ASD-STE100

99

## 7. Cited References

Cortés-Rodríguez, Francisco, 2016: "Parsing Simple Clauses within ARTEMIS: The Computational Treatment of the Layered Structure of the Clause in Role and Reference Grammar", lecture given at the 34th International Conference of AESLA, Alicante.

Cortés-Rodríguez, Francisco, 2019: "ARTEMIS from a constructionist perspective: Reassessing constructional structures for a NLU environment", lecture given at the 8th International Conference on Meaning and Knowledge Representation, Granada.

González-Orta, Marta, and Auxiliadora Martín-Díaz, 2019: "Parsing subordination in ARTEMIS: The case of adverbial complex sentences in STE", paper presented at the 8th International Conference on Meaning and Knowledge Representation, Granada.

Kuhn, Tobias, 2014: "A Survey and Classification of Controlled Natural Languages", *Computational Linguistics* 40 (1), 121-170.

Mairal-Usón, Ricardo, and Francisco Cortés-Rodríguez, 2017: "Automatically Representing TExt Meaning via an Interlingua-based System (ARTEMIS). A further step towards the computational representation of RRG", *Journal of Computer-Assisted Linguistic Research* 1, 61-87.

Mairal-Usón, Ricardo, and Francisco Ruiz de Mendoza, 2008: "Levels of description and explanation in meaning construction" in Chris Butler and Javier Martín-Arista (eds.): *Deconstructing Constructions*, Amsterdam/Philadelphia: John Benjamins, 153-198.

Martín-Díaz, Auxiliadora, 2017: "An account of English yes/no interrogative sentences within ARTEMIS", *Revista de Lenguas para Fines Específicos* (*RLFE*) 23 (2), 41-62.

Martín-Díaz, Auxiliadora, 2018: "Parsing wh-interrogative sentences within ARTEMIS", *Revistælectrónica de Lingüística Aplicada* (*RÆL*) 17 (1), 78-98.

Martín-Díaz, Auxiliadora, 2019: "Parsing adverbial complex sentences in ASD-STE100 within ARTEMIS", *Complutense Journal of English Studies* 27, 159-181.

Periñán-Pascual, Carlos, 2013: "Towards a model of constructional meaning for natural language understanding" in Brian Nolan and Elke Diedrichsen (eds.): *Linking Constructions into Functional Linguistics: The role of constructions in grammar*, Studies in Language Companion Series 145, Amsterdam: John Benjamins, 205-230.

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

Van Valin, Robert, and Randy LaPolla, 1997: *Syntax*, Cambridge: Cambridge University Press.