ARTICLES

Complutense Journal of **English Studies**
ISSN: 2386-3935

EDICIONES COMPLUTENSE

https://dx.doi.org/10.5209/cjes.61076

Parsing adverbial complex sentences in ASD-STE100 within ARTEMIS[1]

María Auxiliadora Martín Díaz[2]

Abstract. One of the main objectives of Natural Language Processing is the simulation of natural language understanding. Within the applications designed for this purpose today, ARTEMIS follows the paradigm of unification grammars (Sag, I., Wasow, T. & Bender, E. 2003), and unlike other trending computational resources, it is theoretically grounded in linguistic models like RRG (Van Valin & LaPolla 1997 and Van Valin 2005), whose linking algorithm lies at the basis of the parsing process of our interlingua-based system.
A fundamental component of ARTEMIS is the Grammar Development Environment (GDE), where feature-based production rules (syntactic, constructional and lexical) are stored and ready to allow the generation of the enhanced layered structure of the clause of natural language expressions (Periñán-Pascual 2013: 222). Syntactic rules for phrasal constituents and simple sentences have already been described (Cortés-Rodríguez 2016; Cortés-Rodríguez & Mairal-Usón 2016; Díaz-Galán & Fumero-Pérez 2015; Fumero-Pérez & Díaz-Galán 2017; Martín-Díaz 2017 and Martín-Díaz 2018), but it now turns to focus on complex sentences, and, to be more precise, on adverbial subordination.
Bearing in mind the validation of these syntactic rules and the common problems that may arise in such parsing applications, our research will concentrate on the analysis of these structures as found in a Controlled Natural Language: ASD-STE100.
Keywords: Adverbial complex sentences, ARTEMIS, syntactic rules.

[es] El proceso de parseado de ARTEMIS para las oraciones adverbiales complejas de ASD-STE100

Resumen. Uno de los principales objetivos del PLN es la simulación de la comprensión del lenguaje natural. Entre las aplicaciones diseñadas para tal fin, ARTEMIS sigue el paradigma de las gramáticas de unificación (Sag, I., Wasow, T. & Bender, E. 2003) y, a diferencia de otros recursos informáticos, se caracteriza por su sólido fundamento teórico. El algoritmo de enlace del modelo lingüístico de la Gramática del Papel y la Referencia (Van Valin & LaPolla 1997 y Van Valin 2005) es central para el proceso de parseado de este sistema basado en la interlengua.
Un componente esencial de nuestro parseador es el Entorno de Desarrollo Gramatical (GDE), en donde se almacenan reglas de producción (sintácticas, contruccionales y léxicas) basadas en rasgos, que son vitales para el proceso de generación de la estructura estratificada de la cláusula enriquecida (Periñán-Pascual 2013: 222). Esta, a su vez, permite a ARTEMIS acometer la interpretación de las expresiones del lenguaje natural. De momento, ya se han descrito las reglas sintácticas del sintagma y la oración simple (Cortés-Rodríguez 2016; Cortés-Rodríguez & Mairal-Usón 2016; Díaz-Galán & Fumero-Pérez 2015; Fumero-Pérez & Díaz-Galán 2017; Martín-Díaz 2017 y Martín-Díaz 2018), pero aún quedan por desarrollar las de la oración compleja.

---

[1] This paper has been developed within the framework of the research project entitled "Desarrollo de un laboratorio virtual para el procesamiento computacional de la lengua desde un paradigma funcional". (UNED) FF2014-53788-C3-1-P funded by the Spanish Ministry of Science.

[2] Department of English and German Philology, University of La Laguna, mmartind@ull.edu.es. Member of the Instituto de Lingüística Andrés Bello (INULAB).

El presente trabajo se centra en el diseño de las reglas sintácticas para las adverbiales complejas en el seno de un lenguaje controlado (ASD-STE100), cuya naturaleza restringida nos ha facilitado la tarea de cara al consiguiente proceso de validación de ARTEMIS.
Palabras clave: Oraciones complejas adverbiales, ARTEMIS, reglas sintácticas.

How to cite this article: Martín Díaz, M. A. (2019) Parsing adverbial complex sentences in ASD-STE100 within ARTEMIS, in *Complutense Journal of English Studies* 27, 159-181.

One of the main objectives of our research project is the validation in the near future of the production rules for the Grammar Development Environment (or GDE module) in ARTEMIS (Automatically Representing TExt Meaning via an Interlingua-Based System) that will derive from our linguistic analysis. To do that, we have opted for studying syntactic 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 evaluate our prototype finally.

In this sense, we have structured this paper by first introducing the type of simplified controlled language to be used for these matters, ASD-STE100, in Section 1. Secondly, a summary of the most critical aspects of complex sentences within Role and Reference Grammar (RRG-Van Valin & LaPolla 1997 and Van Valin 2005) is considered, since ARTEMIS is inspired on this functionally-oriented linguistic theory. In section 3, we can see the influence of the four-level constructional schemata of the Lexical Constructional Model (LCM- Mairal-Usón & Ruiz de Mendoza 2009) on FunGramKB (Functional Grammar Knowledge Base) and ARTEMIS, as computational resources of FUNK Lab , a virtual laboratory for natural language processing where several computational resources (the NLP-LAB, a LABoratory of Natural Language Processing and text analytics; the FunGramKB NAVIGATOR; DEXTER, initials for Discovering and EXtracting TERminology; DAMIEN, initials for DAta MIning ENcountered; and ARTEMIS) with applications in different scientific fields have been built (Mairal-Usón & Cortés-Rodríguez 2017). Section 4 is devoted to the analysis à la ARTEMIS of the adverbial complex sentences identified in ASD-STE100. Finally, to supply the GDE with the necessary tools to carry out a correct parsing of our adverbial complex sentences, a series of Attribute Value Matrixes (AVMs) and production rules (syntactic and constructional) will be proposed to be evaluated in a future research.

1. Controlled languages: ASD-STE100

According to Kuhn, a controlled natural language (CNL) can be defined as "a constructed language that is based on a certain natural language, being more restrictive concerning lexicon, syntax, and/or semantics, while preserving most of its natural properties" (2014: 123). Some controlled languages may be intended to solve communication problems among humans, others to improve manual or machine translation. In particular, ASD-STE100 is a CNL 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. Its initials stand for AeroSpace and Defence Simplified Technical English, but it is often abbreviated to STE or just Simplified English. It had its origins in 1979, even though it did not receive its current name until 2005 when AECMA (*Asociación Española de Construction 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 standard English, but the following restrictive general rules constrain the language at the different levels:

- The lexical level (e.g., "Use approved words from the Dictionary only as the part of speech given")
- The syntactic level (e.g., "You can use the "-ing" form of a verb only as a modifier in a technical name")
- The semantic level (e.g., "Keep to the approved meaning of a word in the Dictionary. Do not use the word with any other meaning."). There is a fixed vocabulary consisting of terms common to the aerospace domain. Additionally, user-defined "Technical Names" and "Technical Verbs" can be introduced.

These restrictions must have necessarily affected the syntax of complex sentences, which would have had to be largely constrained in order to be easily interpreted by parsers like ARTEMIS, especially if the intention behind this is automatic translation.

2. RRG and English Complex Sentences

RRG (Van Valin & LaPolla 1997) accounts for the structure of complex sentences in terms of three types of nexus relation (see Fig. 1 below): Coordination, in which independent structures are related, and two more types in which there seems to be a particular dependency: subordination, where we have dependent structural constituents, and cosubordination where there is operator dependence.

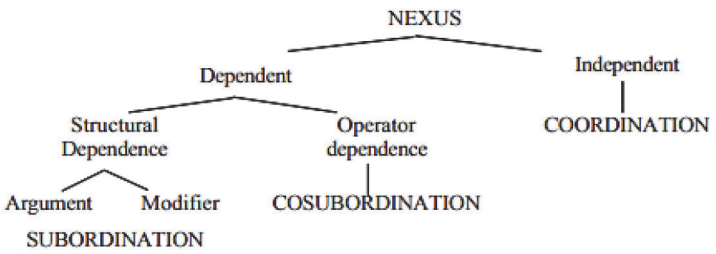

Fig. 1. Nexus types[3].

[3] Taken from VV & LP (1997: 454).

Each of these nexus types can be applied to each of the four levels of juncture that have been identified in RRG: the nuclear level, the core level, the clausal level and the sentential level of juncture. The latter involves the linking of whole sentences and differs from the other linkages in that not all nexus types are possible. According to Van Valin & LaPolla (1997: 469), only coordination is admitted at the sentential level, since cosubordination has no sentential operators to share and subordination has no sentential units to be embedded. However, Van Valin (2005: 192-93) claims that sentential subordination is also possible, as will be explained later on in section 2.4.

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, only nine can be found in English (see below), since nuclear coordination and nuclear subordination have been ruled out by the founders of this grammatical model:

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

A revision, though, concerning a certain ambiguity in the description of some of these subordinate juncts in Van Valin & LaPolla (1997) and Van Valin (2005), needs to be elucidated at this point so as not to render an inconsistent analysis.

Subordination in RRG is characterized by a structural dependence where two types of syntactic constructions can take place: one in which the dependence is in the CORE arguments (daughter subordination), or one in which this dependence is in the modifiers (peripheral subordination, including both relative and adverbial clauses). In Van Valin & LaPolla (1997), adverbial subordination is regarded as a subtype of clausal subordination in which "subordinating conjunctions are treated as predicative prepositions taking a clausal argument and are part of the periphery of the clause" (p. 464). However, this subtype is not considered "a direct daughter of the clause node" since its internal constituents are outside of the "potential focus domain" of the clause (486-87). That is probably why in Van Valin (2005: 194-196), this author distinguishes between two types of peripheral adverbial clauses: temporal clauses which are now classified as a subtype of "[…] 'ad-core subordination' where the subordinate clause is regarded as a modifier of the matrix core and therefore in the $\text{periphery}_{\text{CORE}}$" (p. 194); and other types of adverbials (i.e., reason, concessive, conditional,…), "which express the reason or a condition for the event expressed by the clause as a whole", and consequently, are housed in the $\text{periphery}_{\text{CLAUSE}}$. The rationale for occupying a different layer being that "when the two co-occur in a single sentence, there is a definite preference" for a temporal as opposed to a reason/concessive ordering of the two clauses (p. 195), basically because otherwise, the sentence meaning would be different, as shown in the examples below extracted from Van Valin (2005):

(2) *Kim berated Pat after they arrived at the party because she kissed Chris*.
(3) *Kim berated Pat because she kissed Chris after they arrived at the party* (≠(2))

As we can infer from the explanation above, the difficulty in describing these adverbial juncts within RRG arises from the fact that, even though they are generally considered peripheral, this linguistic model excludes PERIPHERY from the main structural line that goes from the sentence-node to the nucleus. Besides its marginality in the LSC, PERIPHERY is pervasive for RRG and, to a certain extent, the attachment of a given adverbial modifier to a particular layer seems to be open to interpretation even for Van Valin. For the following outline of RRG's complex structures we will follow Van Valin & LaPolla (1997) when classifying peripheral subordination as a clausal juncture (regardless of the subordinator involved), but will adopt Van Valin (2005) when acknowledging the existence of this nexus type at the sentential level, as illustrated in section 3 below for ASD-STE100.

2.1. Nuclear junctures in English

"Nuclear junctures are single cores containing more than one nucleus […] taking a single set of core arguments" (Van Valin & LaPolla 1997: 448). Out of the three possible 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 in English 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 & LaPolla 1995: 446). The two nuclei in these structures permit a variable ordering: they may be adjacent, as in *Push open the door,* 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 (4) below): "the heavier the NP, the more acceptable the examples with adjacent nuclei become" (VanValin & Lapolla 1997: 446).

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

2.2. CORE junctures in English

"In a CORE juncture, […], there is a single CLAUSE containing more than one CORE. Each core may have its own CORE arguments" (Van Valin & LaPolla 1997: 448). All nexus relations (coordination, subordination and cosubordination) can be found in English CORE junctures, where a complementizer or Clause-Linkage Marker (CLM) is usually required to indicate the linked unit in a complex sentence (Van Valin & LaPolla 1997: 470; 2005: 205).

In non-subordinate CORE junctures, the CORES must share an argument, that is, a syntactic and semantic constituent of the matrix CORE which only has a semantic

function in the CORE. Besides, the distinction between coordination and cosubordination lies in the scope of a possible CORE-level operator (i.e. root modality, core directionals, and internal negation (Van Valin & LaPolla 1997: 47).

2.2.1. CORE coordination

In English, CORE coordination is employed for jussive (for commands, requests or demands), direct perception, and propositional attitude (Van Valin & LaPolla 1997: 479 and 481):

(5) *Pat asked the students to leave*
(6) *The king ordered the troops to attack the city*
(7) *Fred made Pam help him*
(8) *Dana saw Chris washing the car*
(9) *Fred saw Harry leave the room*
(10) *Paul considers Carl to be a fool*

In CORE coordination, operators like for example the deontic modal *should* in *Luci should tell Balú to fetch his teddy* only have scope over the first CORE of the layered structure of the clause (LSC), where the control of *should* only governs the CORE *tell Balú*.

2.2.2. CORE Cosubordination

CORE Cosubordination instantiates aspectual, psych-action and purposive relations (Van Valin & LaPolla 1997: 481). Other examples illustrating this juncture-nexus type are grammatical constructions like:

(11) *Kim sat reading*
(12) *Sam sat playing the guitar*
(13) *Carlos must wash the car and clean his room*.

For the last sentence (13), Van Valin & 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*). In this type of CORE cosubordination, CORE nodes are thus dominated by a superordinate CORE node (Van Valin 2005: 203). Therefore, even though in the constituent projection a CORE-level operator like MODD seems to be hanging from the first CORE node, it is clear that its scope must percolate up to the general superordinate CORE and then down to its two daughter COREs, thus allowing the projection of its influence over the NUCs of each CORE node.

2.2.3. CORE Subordination

On the one hand, daughter CORE subordinate constructions are canonically constrained to gerunds and subject *that*-complement clauses, that is, junctures in which

the dependent unit functions as a 'subject' CORE argument of the main clause, as in *That Balú jumped / Balú's jumping upon the bed delighted Luci.* In them, argument sharing is possible (although not obligatory as Van Valin (2005: 189-90) suggests for non-subordinate CORE junctures). This is the case of *Mary regretted slapping Bill the most*, where *slapping*, although syntactically regarded as a subjectless gerund, has a semantic actor in *Mary*. Other examples from Van Valin (2005: 198):

(14) *To wash the car today would be a mistake*
(15) *For John to win the race would be the surprise of the year*
(16) *Washing the car today would be a mistake*

In peripheral subordination, on the other hand, our interest focuses on adverbial clauses functioning as adjunct modifiers[4]. VanValin considers sentences like (17) and (18) below to be Ad-CORE subordinate juncts, where the temporal or spatial setting of the event is expressed through an adjunct adverbial clause that occurs in its periphery (or $\text{periphery}_{\text{CORE}}$). This is so because "the relationship of the adverbial subordinate clause to the core it modifies is the same as that of a peripheral PP modifying a core" (Van Valin 2005: 194), as illustrated in Fig. 2 below.

(17) *Bill went to the party after he talked to Mary*
(18) *John saw Max after he went to the party*

As we can see in this figure, the $\text{periphery}_{\text{CORE}}$ is represented on the margin and attached to the CORE by means of an arrow.

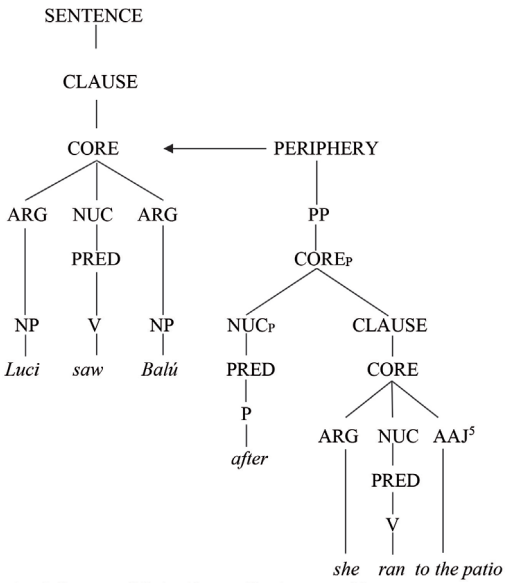

Fig. 2. CORE subordination (adjunct modifier) with a predicative preposition.

[4] Relative clauses will be treated in forthcoming papers.

A second subtype of ad-CORE subordination occurs in English when we have prepositions that have a CORE as their argument. This type of CORE consists of a subjectless gerund, like the ones in the examples from VanValin (2005: 196) reproduced below, and that also displays semantic argument sharing. This subtype of adverbial clause is not available in ASD because, as mentioned in section 1 about the restrictions to this CNL, the only -ing forms allowed in this language are "as a modifier in a technical name".

(19) *Max brushed his teeth after drinking a cup of coffee*
(20) *Chris spoke to his broker before buying more stock*
(21) *Kim threw away the newspaper without reading it*

2.3. CLAUSAL Junctures in English

In clausal junctures, "whole clauses are joined, and each clause may be fully independent of the others" (Van Valin & LaPolla 1997: 448). The core and peripheral constituents of the two clauses are independent because argument sharing does not operate across clause boundaries (Van Valin & LaPolla 1997: 468). In English sentences the three types of clausal juncture-nexus combinations are possible.

2.3.1. CLAUSAL coordination

A defining feature of CLAUSAL coordination (a universal juncture-nexus type), is the impossibility of operator sharing, which means that each clause can have an independent tense, status, evidentiality, and even a distinct illocutionary force (Van Valin & LaPolla 1997: 463-464). For example, in *Sit down and I'll fix you a drink,* the first clause is an imperative and the second an assertion. Both are connected by *and*, a CONJ *in* Van Valin & LaPolla (1997: 464) or a CLM that does not hang from any sentence node in Van Valin (2005: 199). Other examples of CLAUSAL coordination are:

(22) *Anna read for a few minutes, and then she went out*
(23) *Robin is known for liking big parties, but why did she invite the entire club*?

2.3.2. CLAUSAL Cosubordination

At a given level of juncture in cosubordination "[…] the linked units are dependent upon the matrix unit for expression of one or more of the operators for that level." (Van Valin 2005: 201). Clausal cosubordinate juncts in particular exhibit clausal operator dependence, that is, tense, status, evidentiality and illocutionary force must be shared across all juncts. A superordinate CLAUSE-node allows the matrix unit to share both clausal operators between its corresponding cosubordinate daughters, as in the following examples from Van Valin & LaPolla (1997: 455):

(24) *Harry ran down the hall laughing loudly*
(25) *Paul drove to the store and bought some beer*

However, in the following examples of English conjunction reduction, only illocutionary force is shared (Van Valin & LaPolla 1997: 521- 522, Van Valin 2005: 230):

(26) *Kim worked on the assignment in the morning and will finish it in the afternoon*
(27) *Robin drove out of Phoenix this morning and will arrive in Atlanta tomorrow*

2.3.3. CLAUSAL subordination

Subordinate juncts at the level of the clause have no argument sharing, and operator dependence is not significant for them because they are either outside the domain of the IF operator of the clause or have the same force as the main clause (Van Valin & LaPolla 1997: 457). They function either as clausal arguments (clausal subordination daughters) or as clausal adjunct modifiers (clausal adverbial subordinates).

Daughter subordination of the object *that*-complement subtype in English expresses propositional attitude, cognition and indirect relations. As Van Valin (2005:199-200) himself admits this type of juncture is an example of a "syntax-semantics mismatch" that "violates the basic principle that arguments in the logical structure of the verb are realized as core arguments". Accordingly, in a sentence like *Luci decided that she will wash Balú*, the embedded clause *that she will wash Balú* is semantically an argument of the matrix verb *decided*, but syntactically it occurs outside the core, because we can insert a peripheral adjunct like *after school* between the two core elements.

The other subtype of clausal subordination, that of adverbial clauses, is used to indicate for example "the reason or a condition for the event expressed by the clause as a whole" (Van Valin 2005: 194), as is the case of clauses introduced by conjunctions or CLMs like *because*, *if*, *despite* or *although* in English.

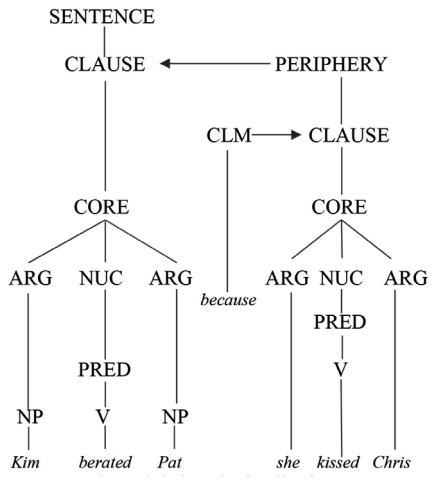

Fig. 3. Ad-clausal subordination.

As we can see in Fig. 3 above, the periphery$_{CLAUSE}$, as well as the periphery$_{CORE}$ in Fig. 2 of section 2.2.3., is represented marginally in the tree scheme. This time, a CLM and not a predicative preposition marks the adjunct clause (Van Valin 2005:

194). This adverbial clause, also on the margin, is introduced by a CLM that links it to the main hierarchical structure using an arrow.

English also has cases of adverbial subordination in the precore slot (PrCs) inside the clause-node, as shown in the example below extracted from Van Valin (2005: 193) *Bill was angry, because after Mary arrived at the party she slapped him* and represented here in the following tree scheme[5]:

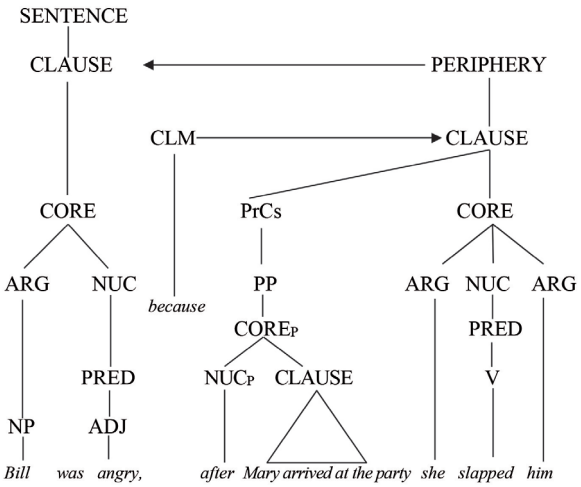

Fig. 4. Ad-clausal subordination.

Examples for this last type of adverbial subordination in the PrCs have not been found for ASD-STE100.

2.4. SENTENTIAL Junctures in English

As mentioned above, sentential junctures involve the linking of whole sentences and, as Van Valin (2005: 191) admits, they are unique in that for them "the full range of nexus types is not available", only sentential coordination and subordination are possible, because cosubordination needs the presence of inexistent sentential operators to be shared.

2.4.1. Sentential coordination

In this type of linkage two complete sentences, or SENTENCE-nodes with their corresponding left-detached positions (LDPs), will be linked by a dominating TEXT-node, as in *As for Sam, Mary saw him last week, and as for Paul, I saw him yesterday*.

---

[5] "The fronted adjunct subordinate clause *after Mary arrived at the party* cannot be in the left-detached position, because it is inside an embedded clause, and embedded clauses cannot in principle have a left-detached position, which is outside of the clause. An embedded clause can, however, have a precore slot, since it is a clause-internal position" (VanValin 2005: 193).

2.4.2. Sentential subordination

According to Van Valin, this type of subordination "involves sentences or clauses occurring in the right- or left-detached positions" (2005: 192). There are two syntactic realizations in English for this type of juncts: the use of direct discourse complements, and the fronting of peripheral adverbial clauses to the left detached position (LDP), a slot outside the clause but within the sentence. The following sentences extracted from VanValin (2005: 192 and 195 respectively) illustrate the second of these syntactic structures.

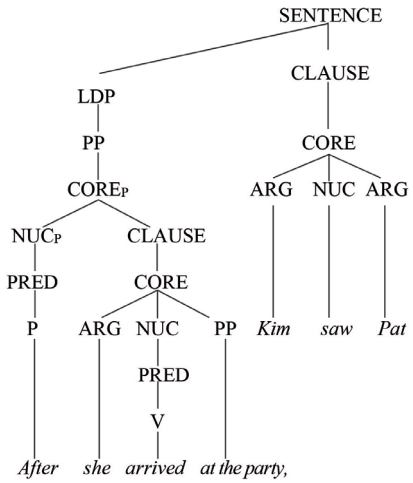

Fig. 5. Sentential subordination (fronting of an original ad-core subordinate).

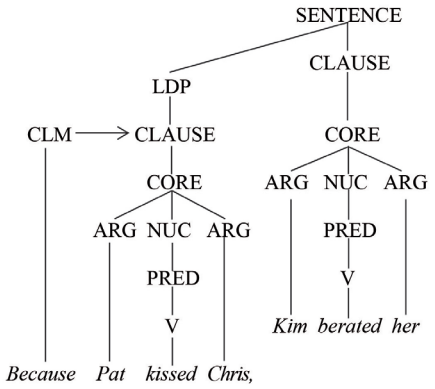

Fig. 6. Sentential subordination (fronting of an original ad-clausal subordinate).

The first of these sentences is introduced by a subordinating conjunction treated as a predicative preposition taking a clausal argument to "express the spatial or temporal setting" of the core event (Van Valin & LaPolla 1997: 464; Van Valin 2005: 194).

The second one begins with a subordinating conjunction that introduces an adverbial modifier to indicate "the reason […] for the event expressed by the clause as a whole" (Van Valin 2005: 194).

However, in their corresponding tree diagrams an LDP lodges both of them, regardless of the level at which the linkage originally occurred and the syntactic and/or semantic difference lying between their initial components. This fact could indicate that a more similar syntactic analysis in ARTEMIS could perhaps facilitate both their parsing and their semantic interpretation. In line with this, Van Valin & LaPolla themselves claim in their Interclausal Relations Hierarchy, that "the burden of expressing the semantic relations" among the units of a complex structure really "falls on the clause-linkage markers" (1997: 477).

Subjectless cores similar to those in section 2.2.3 for core subordination could also participate in these sentential junctures (*After arriving, he ...*). For obvious reasons, STE has no examples for this complementation.

3. Artemis and the Parsing of Adverbial Subordinates

Within FUNK-Lab's resources, ARTEMIS is a prototype application designed within the paradigm of constraint-based grammars to enable the understanding of natural languages in the framework of RRG. FunGramKB is, in turn, a knowledge base which provides ARTEMIS with a "large-scale repository of fine-grained morpho-syntactic, semantic and pragmatic knowledge" on which to base an effective parsing (Periñán-Pascual & Arcas-Túnez 2014: 181).

Apart from counting on powerful tools to encode natural-language sentences into a machine-readable expression, such as the CLS constructor[6] and the COREL-scheme Builder[7], ARTEMIS consists of a third module, the Grammar Development Environment (or GDE), where the grammar building process takes place. In it, lexical, syntactic and constructional information of specific languages is stored by means of two components: an inventory of production rules and a catalogue of Attribute-Value Matrixes (AVMs), i.e., "Complex formal descriptions of grammatical units" that constrain the parsing process (Periñán-Pascual 2013).

Syntactic and constructional rules in ARTEMIS are intended to computationally enrich the framework of RRG's Layered structure of the Clause (LSC) and its linking algorithm. Within construction rules, a clear distinction between kernel and non-kernel level-1 constructions (L1-constructions) is established whereby the kernel constructions (for example, the monotransitive kernel-2 construction *John kicked the ball*) are derived from the Lexicon, and the non-kernel constructions (for example, the L1-transitive-resultative construction *John kicked the ball flat*, or the L1-caused motion construction *John kicked the ball into the stadium*) are generated, even recursively (for example, in *John kicked the ball flat into the stadium*) with the aid of the core grammar of the verb together with all its constructional schemata (Periñán-Pascual 2013: 214). This constructionist linguistic view was promoted in ARTEMIS by the LCM, thanks to which we can store in FunGramKB's Grammati-

---

6 A tool for the generation of a Conceptual Logical Structure.

7 A tool to transform a CLS into a COnceptual REpresentation Language and make ARTEMIS useful for NLP tasks.

con four different types of constructional meaning (argumental, or L1; implicational, or L2; illocutionary, or L3; and discursive, or L4) by way of multilevel constructional templates.

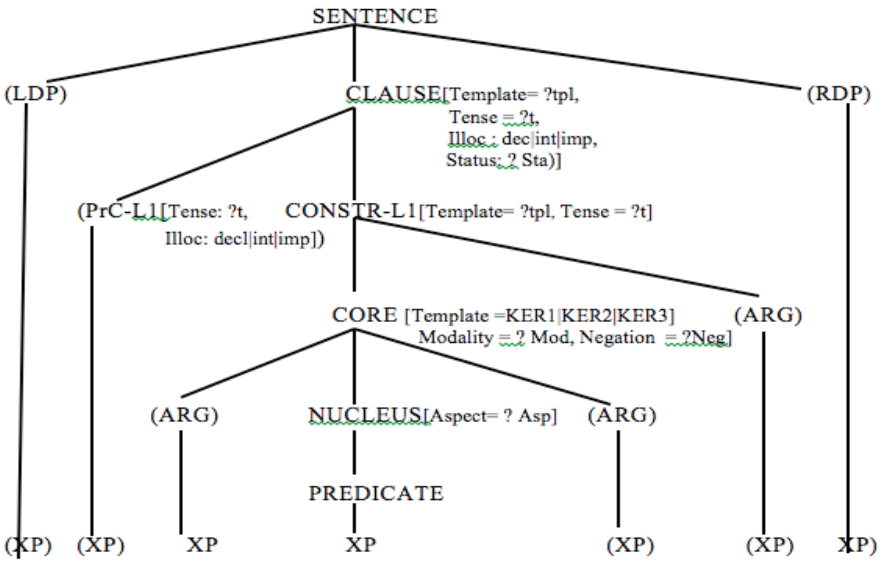

Fig. 7. Modified LSC in ARTEMIS.

This kernel/non-kernel distinction involved two important modifications for RRG's LSC: firstly, the introduction of a new constituent in this hierarchical structure, the CONSTR-L1 node; and secondly, the subsequent redefinition of RRG's Precore slot position as a Preconstruction-L1-position, the PrC-L1 node "where constituents triggered by a construction can also intervene" (Cortés-Rodríguez & Mairal-Usón (2016). Both adjustments are shown in Fig. 7 above.

The idea that constructional meaning can improve the descriptive capacity of a semantic theory led Periñan-Pascual & Arcas-Túnez (2014) to define construction as

> a pairing of form and meaning, serving as a building block in the compositionality of sentential semantics, whose meaning cannot be fully derived from the sum of the lexical meanings of the individual constructs taking part in the utterance (172)

and therefore, to integrate it as a universal category into RRG's LSC (Periñan-Pascual & Arcas-Túnez 2014:171). This CONSTR-L1 node for L1-constructions halfway between the CORE and the CLAUSE nodes basically derives from either the inclusion of a secondary nucleus (NUC-S) or an argument adjunct (AAJ). Besides, its integration implied that the clause was now seen as a layer configured "as one or more L1-constructions which are recursively arranged" and where "the innermost construction introduces the core, which can be modelled by other L1-constructions, typically contributing with a further argument" (Periñán-Pascual 2013: 222). This modification also implied the redistribution of original peripheral modifiers in Van

Valin (2005) to its own periphery$_{\text{L1-CONSTRUCTION}}$, as shown below in Fig. 8 (taken from Periñán-Pascual 2013: 221).

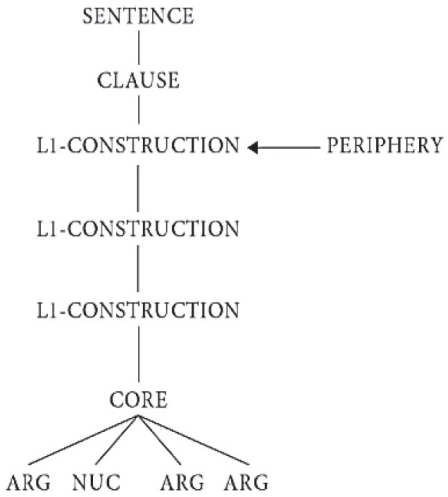

Fig. 8. Enhanced model of LSC (refined tree).

In line with these modifications, a tree-diagram of an adverbial subordination is presented in Fig. 9 below. It is not possible to account for this syntactic realization in ARTEMIS without resorting to our catalogue of constructions (FunGramKB's grammaticon) and searching for the linguistic description (i.e., the constraining AVMs) of its corresponding constructional components, an L1-resultative and an L1-caused motion constructions (Mairal-Usón & Cortés Rodríguez 2017). Unlike RRG, where lexical elements like *flat* and *out of the stadium* are syntactically represented in their respective tree-structures as a daughter-NUC of a more complex and superordinate NUC, and as an AAJ, FunGramKB adopts a more constructional approach where machine-readable devices like the above-mentioned AVMs can also store constructional meaning ready to be used by a parsing application like ARTEMIS. This approach means that in our example below an AVM for the resultative construction will have to unify with the lexical entry of the predicate *kick* to give place to the first L1-CONSTR. This output, in turn, will have to unify with another AVM for the caused-motion construction, thus producing the second L1-CONSTR.

An adverbial clause further complements the sentence in Fig. 9 above in the periphery, whose optional attachment as an adjunct is represented using an arrow in RRG. This representation poses a problem for ARTEMIS, firstly, because we can infer from it that its optionality is not as relevant as to deserve the slot a structural constituent does. Secondly, because in RRG "the linear order of the core arguments and the predicate is irrelevant to the determination of whether an element is in the nucleus, core or periphery" (Van Valin & LaPolla 1997: 32). As a computational parsing application, ARTEMIS must follow, however, a linearity of processing so that a tag or label can be assigned to each of the constituents in the sentence, and analyze them in a strict sequential order.

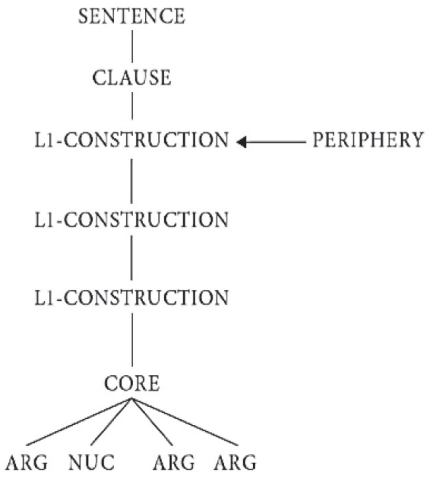

Fig. 9. Tree-diagram of an adverbial modification of a caused-motion and resultative L1-construction.

Moreover, on linking syntax to semantics in complex sentences, Van Valin (2005), as opposed to Van Valin & LaPolla (1997), establishes an interclausal relations hierarchy asserting that if the interclausal semantic relation is circumstantial, that is, it concerns spatial or temporal parameters of an event, "the predicative preposition plus its object, be it an NP, a core or a clause, is an ad-core modifier". On the other hand, if the semantic relation involves reason, condition or concession, "the resulting adjunct PP or clause is linked to the periphery$_{\text{CLAUSE}}$" (chapter 7). RRG syntactic analysis is "based on an inventory of templates […] which do not explicitly state the order of constituents but just their hierarchical organization" (Periñán-Pascual 2013: 222), so, surprisingly, Van Valin seems to substantiate the different hierarchical position of temporal over reason clauses in a preferential linear precedence (see section 2 above), which, to me, does not necessarily correlate with a distinct layering. In fact, Quirk et al. (1985: 1042) admit that if a complex sentence comprises two final subordinate clauses, generally the "final subordinate clause is interpreted as subordinate to the immediately preceding clause".

More in line with the type of formalism we need for a parsing application like ARTEMIS based on unification grammars, the taxonomy of RRG adjuncts proposed by Díaz-Jorge (2017) comes to support the view that both temporal and contingency adjuncts (within which we can include reason, purpose and concession, following Quirk et al. (1985: 564)) must be attached to a single node, the periphery$_{\text{CORE}}$, leaving for the periphery$_{\text{CLAUSE}}$ epistemic, illocutionary and evidential adjuncts, and reducing to a preferential order their sequencing within the core.

These considerations above make us conclude that both types of peripheral juncts can be best described as belonging to a single layer, to which them as modifiers of the output of the last L1-construction can contribute (cf. figures 8 and 9 above). Furthermore, the computational requirement of designing feature-based production rules in the GDE which must be subject to the order or linearization of their constituents implies having to re-interpret this periphery, because of its optionality, as a

latent daughter node of the clause, only taking part in it when an L4-cause construction, in this case, occurs in the discourse.

Likewise, and for the sake of facilitating the linear parsing process within ARTEMIS, a rather liberal interpretation of the label CLM is introduced, as discussed in section 2 above. These adjustments will include as possible lexical realizations of this CLM, not only the adverbial subordinators categorized as such in Van Valin (2005) (*because, if, although*, etc.), but also temporal subordinators like *after* and *before* (interpreted as predicative prepositions in RRG), and further types of CLMs that can trigger other types of juncture-nexus combinations (i.e., *to, from, and, that*, etc.). This CLM-node reinterpreted now as a necessary constituent in the LSC will also need an AVM to be associated with in the parsing process (see (28) below). As possible attributes of this new category we have two candidates: the attributes of NEXUS, and "syn", for the syntactic co-occurrence that may be produced, for example, in juncture nexus combinations where subjectless gerunds must follow[8]. Whereas the attribute "syn" has already been developed in Cortés-Rodríguez and Mairal-Usón (2016), a subsequent AVM for the attribute NEXUS is also required to encode the values of coordination, cosubordination and subordination.

(28) <Category type= "CLM">
<Attribute ID="Nexus"/>
<Attribute ID="Syn"/>
</Category>

(29) <Attribute ID="Nexus" obl="+" num="1">
<Value>?nexus</Value>
<ValueTag="coordinating">co</Value>
<ValueTag="cosubordinating">csb</Value>
<ValueTag="subordinating">sb</Value>
</Attribute>

The attribute nexus in the CLM is lexicalized by a specific conjunction that will enable the triggering of a given complex template from FunGramKb's grammaticon, thus allowing ARTEMIS to unify this syntactic structure with a specific type of L4-construction.

As for the syntactic rules to be stored in the GDE for our adverbial complex sentences, we can conclude that only two types of juncture-nexus combinations are distinguished, both of them introduced by a CLM conjunction[9]: i. the clausal subordination type; and ii. the sentential subordination type. In the first case, two subtypes of clausal subordinates are generally observed in English (30a and 30b below). In the case of (30a) the clause is modified by an adverbial clause in the periphery which is introduced by the CLM *because*. In (30b) the matrix clause is again modified by a peripheral subordinate (introduced as well by *because*), but further modified by another adverbial clause in the PreC-L1 position (introduced this time by *after*). Out of these, only the first of these syntactic rules works in our CNL (see section 4).

---

8 a given CLM See examples (19), (20) and (21) in section 2.2.3 above.

9 For other complex sentences, a CLM preposition (*to* or *from*) will be proposed (Van Valin & LaPolla 1997: 8.4.2).

(30) a. *Kim berated Pat because she kissed Chris.*
CLAUSE → CONSTR-L1 – PERIPHERY, where
PERIPHERY → CLAUSE, where
CLAUSE → CLM – CONSTR-L1

b. *Bill was angry, because after Mary arrived at the party she slapped him.*
CLAUSE → CONSTR-L1 – PERIPHERY, where
PERIPHERY → CLAUSE, where
CLAUSE → CLM – PrC-L1– CONSTR-L1, where
PreC-L1 → CLAUSE, where
CLAUSE → CLM – CONSTR-L1

On the other hand, only one type of adverbial subordination at sentential level has been described in English (see (31) below). The sentence layer of this example is concerned because the adverbial modification has been fronted to an extraclausal position, the LDP, where a CLM *after* introduces the subordinate clause. Examples of these syntactic rules abound in ASD-STE100 (see section 4).

(31) *After Kim arrived at the party, Pat saw her.*
SENTENCE → LDP – CLAUSE, where
LDP → CLAUSE, where
CLAUSE → CLM – CONSTR-L1

4. Parsing Adverbial Subordinates of ASD-STE100

One of the most frequent juncture-nexus combinations found in ASD-STE100 for adverbial subordination is that developed in the periphery of the clause and introduced by different CLMs.

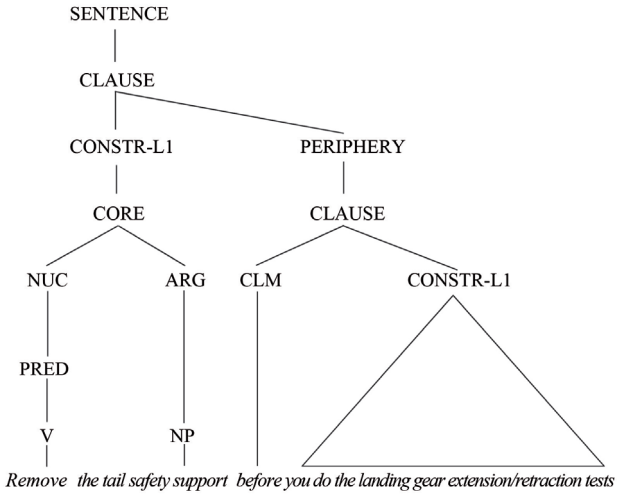

Fig. 10. Temporal Clause subordination in STE.

This junct subtype mostly functions as a temporal modifier (Fig. 10 above), a reason modifier (Fig. 11 below) or conditional modifier (Fig. 12 below) of the matrix core. Additional examples for temporal sentences with different CLMs in STE[10] are:

(32) *Do a functional test **after** you install the component.*
(33) *Do this **until** you see no air bubbles in the fluid.*
(34) *Do not bend the electrical harness too much **when** you release or tighten the clips or clamps.*
(35) *Let the brakes and the wheels become cool **before** you go near the landing gear.*
(36) *Hold the upper and lower torque links **while** you remove or install the apex pin.*

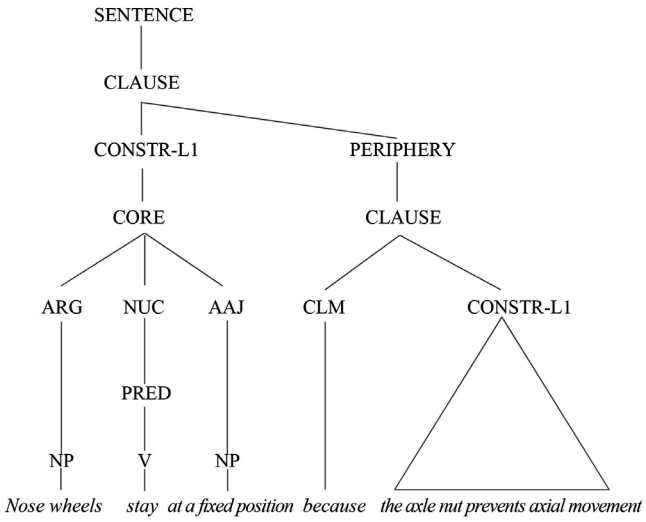

Fig. 11. Reason Clause subordination in STE.

Only one case with the conjunction *so* (*The Parking Brake System does not have a differential braking function, **so** the system does not adjust the supplied brake hydraulic-pressure*), presumably not accepted in STE, has been found[11].

The other subtype of juncture nexus combination found for adverbial complex sentences in STE is Sentential Subordination, characterized by fronting adverbial clauses to the extraclausal LDP. The following figures illustrate temporal, concessive, conditional and reason sentences.

[10] In those cases in which the CLM is *when*, an alternative analysis could have been to interpret the given adverbial clause as having a PrC-L1 node saturated by this Wh-form. However, for the sake of abridging the parsing in ARTEMIS and making it more similar to that of other complex sentences with other temporal CLMs (see Fig. 10 above), a conjunction reading of *when* has been preferred.

[11] According to Quirk et al. (1985), *so* is a contingency adjunct and, within these, a result adjunct.

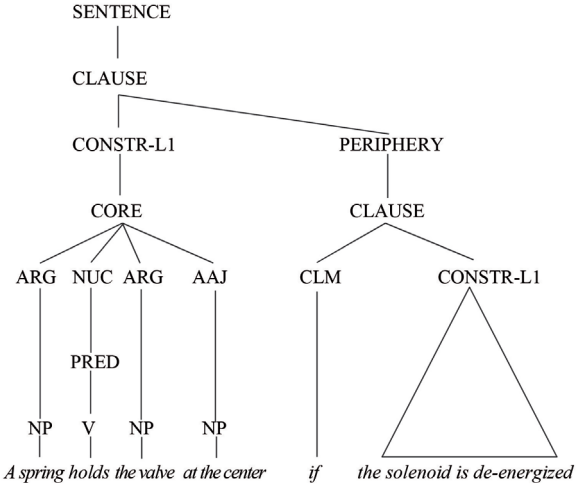

Fig. 12. Conditional clause subordination in STE.

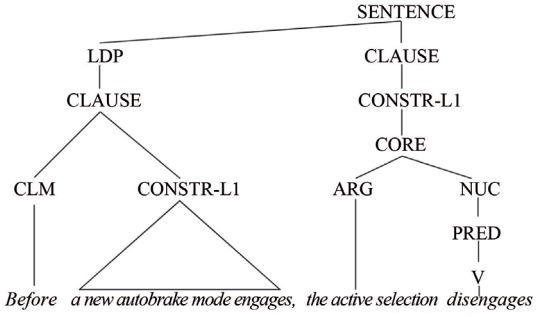

Fig. 13. Temporal sentence subordination in STE.

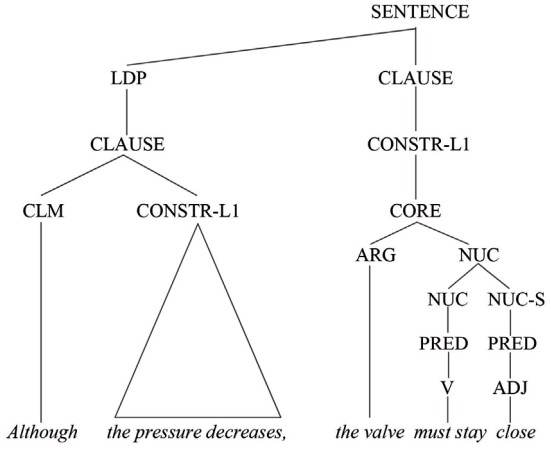

Fig. 14. Concessive sentential subordination in STE.

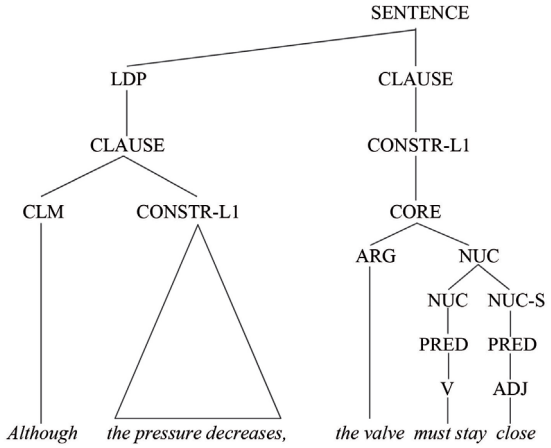

Fig. 15. Conditional sentence subordination in STE.

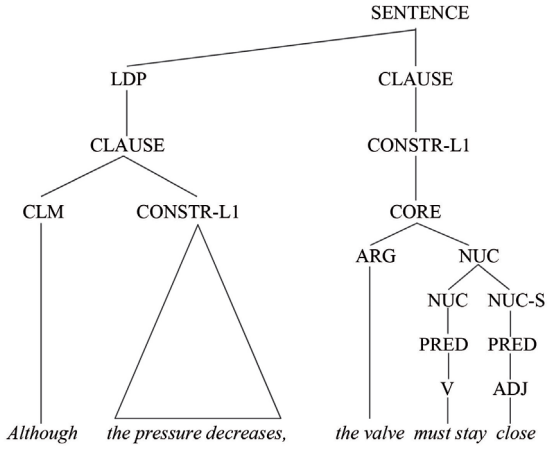

Fig. 16. Reason sentence subordination in STE.

Other examples for this type of ad-sentential junct with different subordinators in STE:

(37) *And, **as** the rotors are connected to the wheels, the rotors decrease the wheel speed*
(38) ***After** the engines stop, none of the two functions, antiskid or differential braking, are available from the brake-accumulator pressure-source*
(40) ***When** the fluid becomes stable in the container, it shows its normal colour*
(41) ***While** you remove the pin, hold the spacer and torque link*

5. Conclusion

The analysis à la ARTEMIS of adverbial complex sentences in ASD-STE100 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 the establishment of two basic types of adverbial structures in English: i. peripheral adverbials within the CLAUSE; ii. fronted adverbials at SENTENCE level. Both of them have been found in the analysis we have carried out in our aeronautical controlled language, even though not with the same diversity as that generally seen in English. As expected from a CNL, complex structures like that in 30b above were not encountered in ASD-STE100.

The most important of the adjustments proposed here has implied the redefinition of a marginal and multilevel PERIPHERY in RRG into a structural daughter-node of the CLAUSE that can be optionally saturated in the parse tree of ARTEMIS. In unmarked situations, that is, in the case of kernel constructions, such a PERIPHERY would modify the CORE; whereas in marked situations, like that of an L1-Construction, this peripheral component would involve the CONSTR-L1 node. A second modification arises from the proposal in the present paper of a CLM conjunction as an essential constituent of both peripheral and fronted adverbial subordinates. This fact has allowed us to design syntactic rules for the GDE where the CLM becomes the initiator of a subordinating clause in the PERIPHERY, the PreC-L1 position (although examples for this one have not been found in ASD-STE100), or the LDP-node, all of them at the edges of the LSC. This circumstance could facilitate ARTEMIS the task of filtering out the relatively high percentage of complex syntactic structures that may be initiated in English by a certain CLM. Within the attributes suggested for the AVM of the category CLM, "nexus" becomes relevant to account for the interaction between the syntactic rules for the adverbial sentences here advanced and the L4-constructional templates in FunGramKB's Grammaticon.

Testing the prototype is fundamental to confirm that the parser works as expected, therefore the following step in this direction will be to complete the endeavour of accounting for the effect of constructional interaction with the theory of nexus in this and other complex sentences in ASD-STE100, and evaluate its success rate after comparing it to similar computational resources.

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 RRG: 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. (2016). Developing Parsing Rules within ARTEMIS: The case of Do Auxiliary Insertion. In C. Periñán-Pascual & E. Mestre Mestre (Eds.), *Understanding meaning and knowledge representation: from theoretical and cognitive*

linguistics to natural language processing (pp. 283-302). Cambridge: Cambridge Scholars Publishing.

Díaz-Jorge, V. (2017). *Adjuncts in role and reference grammar: the peripheries*. Trabajo de fin de Grado dirigido por el profesor Francisco J. Cortés Rodríguez en el Grado de Estudios Ingleses del Departamento de Filología Inglesa de la Universidad de La Laguna.

Fumero-Pérez, M. & Díaz-Galán A. (2017). The Interaction of parsing rules and argument- predicate constructions: implications for the structure of the Grammaticon in FunGramKB. *Revista de lingüística y lenguas aplicadas,* 12: 33-44.

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

Mairal-Usón, R. & Cortés Rodríguez, F. (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.

Martín-Díaz, M.A. (2017). An account of English yes/no interrogative sentences within ARTEMIS. *Revista de lenguas para fines específicos (RLFE),* 23(2): 4-62.

Martín-Díaz, M.A. (2018). Parsing wh-interrogative sentences within ARTEMIS. *Revista electrónica de lingüística aplicada (RAEL)* 17 (1): 78-98.

Periñán-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.

Periñán-Pascual, C. & Arcas-Túnez F. (2014). The implementation of the CLS constructor in ARTEMIS. In B. Nolan & C. Periñán-Pascual (Eds.), *Language processing and grammars: the role of functionally oriented computational models* [Studies in Language Companion Series 150] (pp. 165-196). Amsterdam: John Benjamins.

Quirk, R., Greenbaum S., Leech, G. & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.

Sag, I, Wasow, T. & Bender, E. (2003). *Syntactic theory: formal introduction*. Stanford: CSLI Publications.

Van Valin, R. (2005). *Exploring the syntax-semantics interface*. Cambridge: Cambridge University Press.

Van Valin, R. & LaPolla, R. (1997). *Syntax*. Cambridge: Cambridge University Press.

Appendix: List of abbreviations

AAJ Argument-adjunct
ADJ Adjective
ADVR Relative adverb
ARG Argument
AUX Auxiliary verb
AVM Attribute-Value Matrix
CL Clause
CLM Clause Linkage Marker
CLS Conceptual Logical Structure
CNL Controlled Natural Language
CONSTR-L1 Level 1 Construction
GDE Grammar Development Environment

LCM Lexical Constructional Model
LDP Left detached Position
LSC Layered Structure of the Clause
MODD Modal verb (deontic)
MODST Modal verb (epistemic)
N Noun
NUC Nucleus
NUC-S Secondary Nucleus
PER Periphery
PP Prepositional Phrase
PrCS PreCore Slot
PreC-L1 Pre L1 Construction Slot
PRED Predicate
RRG Role and Reference Grammar
S Sentence
STE Simplified Technical English