Computational analysis of adjuncts in ASD-STE100 for the NLP parser ARTEMIS

Francisco José Cortés-Rodríguez
"Andrés Bello" Institute of Linguistics
Universidad de La Laguna
fcortes@ull.edu.es

Carolina Rodríguez-Juárez
Research Institute of Text Analysis and Applications
Universidad de Las Palmas de Gran Canaria
carolina.rodriguez@ulpgc.es

Abstract

Texts written in Simplified Technical English present semantic and syntactic restrictions with respect to standard English that should be considered in the implementation of natural language processing systems dealing with controlled natural languages. This paper explores the syntax and semantics of adjuncts in a controlled natural language, namely, the Simplified Technical English (ASD-STE100) used in the Airbus corpus with a view to observing how the peculiarities identified in the behaviour of Airbus adverbials should be reflected in the computational treatment of these constituents. Thus, our main aim is to contribute to the computational implementation of the prototype ARTEMIS by designing the parsing rules and the catalogue of feature-bearing matrixes that encode the grammatical constraints of Airbus adjuncts. The parsing rules reflect the positional preferences of the various peripheral types and capture the semantic variability of adjuncts in the corpus. In addition, they provide a weight factor that predicts the scale of markedness of these constituents with respect to the different positions that they occupy in the clause. On the whole, these properties offer a precise description of the syntactic features of adverbials that will facilitate their automatic processing.

Keywords: adjuncts, Airbus corpus, ARTEMIS, ASD-STE100, parsing rules.

Resumen

Los documentos escritos en Inglés Técnico Simplificado (ASD-STE100) muestran determinadas restricciones semánticas y sintácticas que deben tenerse en cuenta para la creación de sistemas de procesamiento de lenguaje natural aplicados a lenguajes controlados naturales. En este trabajo se exploran los rasgos sintácticos y semánticos de los adverbiales en uno de estos lenguajes controlados naturales, el ASD-STE100, utilizado en el corpus Airbus. El propósito de este análisis es establecer el modo en que las peculiaridades del comportamiento de los adverbiales en este corpus deben encontrar reflejo en el tratamiento computacional de dichos constituyentes. Con ello pretendemos contribuir a la implementación del prototipo ARTEMIS mediante el diseño de las reglas para el análisis sintáctico (reglas de parseado) y las matrices atributo-valor de este prototipo, las cuales codifican las restricciones gramaticales de los adjuntos del corpus de Airbus. Las reglas de parseado reflejan las preferencias posicionales de los diferentes tipos de periferias adverbiales y capturan la variabilidad semántica de los adjuntos en el corpus. Además, ofrecen un factor de peso que predice la escala de marcado de estos constituyentes con respecto a las diferentes posiciones que ocupan en la cláusula. Estas propiedades ofrecen en conjunto una descripción precisa de los rasgos sintácticos de los adverbiales que facilitará el procesamiento automático de los mismos.

Palabras clave: adjuntos, ASD-STE100, ARTEMIS, corpus Airbus, reglas de parseado.

1. Introduction

The complexity of natural languages presents a problem for computers when trying to process and understand them due to the ambiguity and the implicit meaning that these languages present. Formal languages emerged as a possible solution to this issue but have turned out to be difficult to understand by domain specialists as they "cause a cognitive distance to the application domain that is not inherent in natural language" (Schwitter, 2010: 1113). Thus, an intermediate solution between these two types of languages is the use of controlled natural languages (CNLs) since they have been designed to reduce the ambiguity and complexity of natural languages (Schwitter, 2010: 1113). These constrained natural languages belong to different environments and disciplines, and are constructed in accordance with "a well-defined subset of a language's grammar and lexicon" (Kittredge, 2003, quoted in Kuhn, 2014: 122). Besides, they include the specific technical vocabulary needed in a particular domain.

In 2010, Schwitter claimed that it was "an exciting time to work on controlled natural languages" (p. 1120). In fact, over the past years, numerous studies have focused on the development of machine-oriented controlled natural languages (e.g. Attempto Controlled English (ACE, Fuchs et al., 2008), Processable English (PENG, White & Schwitter, 2009) or Computer Processable Language (CPL, Clark et al., 2010), to name a few) that are aimed to "improve the translatability of technical documents (e.g. machine translation (...)) and the acquisition, representation, and processing of knowledge (e.g. knowledge systems (...)) and in particular for the Semantic Web"

(Schwitter, 2010: 1114). In this paper, however, we focus on another category of CNLs, namely, human-oriented CNLs whose aim is to facilitate the comprehensibility and the readability of technical texts (Schwitter, 2010: 1113), such as the ones contained in the Airbus corpus. The Airbus corpus is made up of maintenance documents written according to the standards and rules specified in the document "ASD Simplified Technical English Specification (ASD-STE100)" (2017). The need to develop this simplified language (STE, Simplified Technical English) emerged in the field of the aerospace and defense industry since the writers of these technical documents had to guarantee that readers, mainly airlines staff (of whom 80% are non-native speakers of English), would be able to understand maintenance and operation documents to guarantee the aircraft availability without putting human lives at risk (https://asd-ste100.org/about.html). Such has been the success of STE that industries from areas as diverse as language services, professional translation and interpreting and the academic world are also using this STE (https://asd-ste100.org/index.html).

The issues of ASD-STE100 are constantly being updated to catch up with the technological evolution,[1] but its structure is stable and consolidated, with a set of Writing Rules (Part 1) covering aspects of grammar and style, and a Dictionary of controlled vocabulary (Part 2) that lists the words that are approved and, as a result, can be used (https://asd-ste100.org/about.html). Just to illustrate why this technical language is regarded as a "simplified" language, let's show a few representative examples taken both from the ASD-STE100 specification document and from the Airbus corpus. ASD-STE100 (2021) restricts the parts of speech of a particular word that can be used. Thus, the word "test" can only be used as a noun and not as a verb:

1. STE: <u>Test</u> B is an alternative to <u>test</u> A.

   Non-STE: <u>Test</u> the system for leaks.

   STE: Do the leak <u>test</u> of the system / Do a <u>test</u> for leaks in the system.

   Airbus corpus: Wait for a minimum of 2 seconds before you launch the <u>test</u>.

   Phrasal verbs with idiomatic/abstract meanings cannot be used:

2. NonSTE: This compound can <u>give off</u> poisonous fumes.

   STE: This compound can <u>release</u> poisonous fumes.

   Airbus corpus: The forward kneeling manifolds open to <u>release</u> hydraulic flow ...

1 The last updated version of ASD-STE100 dates back to April 30, 2021, ISSUE 8 (https://asd-ste100.org/).

STE recommends not omitting verbs or subjects, because the reader will not understand what the action is or what you are referring to:

3. Non-STE: Rotary switch to INPUT.

   STE: Set the rotary switch to INPUT.

   Airbus corpus: Put the rotary switch (5) in position on control panel 11VU (3).

4. Non-STE: If installed, remove the shims.

   STE: IF shims are installed, remove them.

In procedural writing, for example, ASD-STE100 recommends writing short sentences with a maximum of 20 words, whereas in descriptive writing a maximum of 25 words are allowed. In instructions, the imperative form must be used and only one instruction should be included in one sentence:

5. Non-STE: Set the TEST switch to the middle position and release the SHORT-CIRCUIT TEST switch.

   STE: A. Set the TEST switch to the middle position.

   B. Release the SHORT-CIRCUIT TEST switch.

6. Airbus corpus:

   Release the MLG 1M electrical-harness (21) from the upper bracket.

   Remove the upper bracket (22) from the kneeling cylinder (14).

Natural Language Processing (NLP) systems need to be adapted to the semantics and syntax of CNLs for a successful automated natural-language processing. Thus, with this research we aim to contribute to the development of an NLP prototype called ARTEMIS (Automatically Representing Text Meaning via an Interlingua-based System) (Periñán-Pascual, 2013a/b; Periñán-Pascual & Arcas-Túnez, 2014) that has been designed to obtain the syntactic and semantic representation of linguistic structures and that has been implemented as a parsing device within the lexico-conceptual knowledge base FunGramKB (Periñán-Pascual, 2012, 2013a/b; Periñán-Pascual & Arcas-Túnez, 2014). At the moment, the ARTEMIS parser is being bench tested for the controlled natural language ASD-STE100 with the idea of achieving the parsing of standard English eventually (Fumero-Pérez & Díaz-Galán, 2019: 152). In fact, this paper complements previous studies focused on the implementation of the computational grammar of ARTEMIS through the development of the production rules (lexical, syntactic and constructional) that are stored in the Grammar Development Environment (GDE) and that take part in the parsing process of linguistic expressions. In particular, these studies have addressed the formalised description of constructional and non-constructional meaning in ASD-STE100, and the present research on ASD-STE100 adjuncts contributes to the further development of the GDE by exploring the semantic and positional variability of adverbs in this CNL.[2]

In this article, we aim to enrich the investigation carried out so far by designing the Airbus syntactic rules for the internal realization of adverbials as well as for the localization of the peripheries at the different layers of the constituent structure of the clause. These rules will reflect the notable mobility and semantic variability of these components, aspects that have been registered and analysed in a previous quantitative corpus-based study (Rodríguez-Juárez & Cortés-Rodríguez, forthcoming). We also aim to provide the set of Attribute-Value Matrixes (AVMs) for ASD-STE100 adverbial units that are configurated as a list of descriptors (attributes) and values that establish the semantic restrictions and/or selection constraints that cannot be directly retrieved from the modules of the knowledge base FunGramKB in which the parser is implemented; in particular, they cannot be obtained from the Lexicon (*Aktionsart* ascription, macrorole assignment, variables, logical structures, etc.), the Grammaticon (inventory of grammatical constructions) or the Ontology (hierarchy of conceptual units). These AVMs have additionally been enriched with the addition of the attribute "weight" that predicts the scale of markedness of nuclear and core adjuncts per position in the clause, thus offering a precise description of their syntactic representation that will facilitate the automatic processing of Airbus adjuncts in NLP tasks.

The remainder of the paper is structured as follows. Section 2 offers a brief overview of the parser ARTEMIS and of the lexico-conceptual knowledge base FunGramKB in which it is implemented. It additionally presents a general account of the grammatical models on which FunGramKB and ARTEMIS are grounded: Role and Reference Grammar and the Lexical Constructional Model. In section 3, we describe the corpus and the methodology followed in this research. In section 4, we

2 Examples of the investigation carried out so far into the design of the production rules for ARTEMIS and, more specifically, into the formalised description of ASD-STE100 are the following: complex grammatical structures withing the Nucleus layer (Cortés-Rodríguez & Mairal-Usón, 2016), DO-auxiliary insertion (Díaz-Galán & Fumero-Pérez, 2016), *yes-no* interrogative sentences (Martín-Díaz, 2017), simple clauses (Díaz-Galán, 2018), *wh*-interrogative sentences (Martín-Díaz, 2018), phrasal constituents (Cortés-Rodríguez & Rodríguez-Juárez, 2018), function words (Fumero-Pérez & Díaz-Galán, 2019), adverbials (Cortés-Rodríguez & Rodríguez-Juárez, 2019), adverbial complex sentences (Martín-Díaz, 2019), non-propositional meaning (Díaz-Galán & Fumero-Pérez, 2020), subordinate clauses (Martín-Díaz & González-Orta, 2020), non-peripheral complex sentences (González-Orta & Martín-Díaz, 2022).

include our main contributions by propounding the catalogue of AVMs for adjuncts and nuclear and core modifiers, as well as the attributes for "AdjunctRole", "concept" and "weight" (subsection 4.1), and by outlining the rules for the internal realization of peripheries (adjuncts, nuclear peripheries and level-1 peripheries) (subsection 4.2), and for the localization of peripheries in the layers of the layered structure of the clause (subsection 4.3). Section 5 wraps up the general contributions and conclusions of our research.

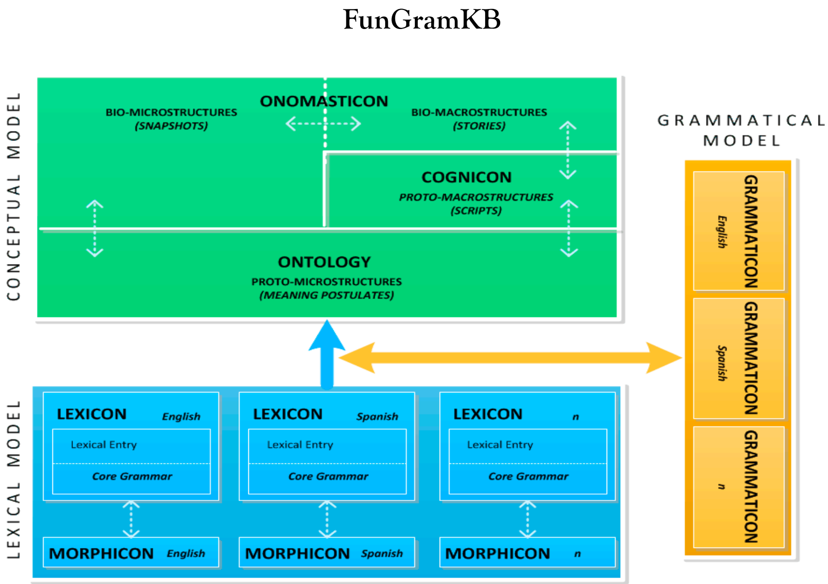
2. Background for the analysis: an overview of ARTEMIS and FunGramKB

Since the final aim of this research is to provide the production rules and Attribute-Value Matrixes (AVMs) that are needed for the computational parsing of Airbus adjuncts (also referred to as peripheral constituents in the remainder of this paper) in ARTEMIS in sentences like "*Lightly* loosen the axle nut (6)" or "*After 15 minutes*, do a leak check of these areas", we need to offer a brief overview of the parsing device ARTEMIS and of the knowledge base FunGramKB from which the NLP prototype extracts the relevant information for the effective semantic and syntactic parsing of sentences. FunGramKB and ARTEMIS are framed within the grammatical theory of Role and Reference Grammar (RRG) (Van Valin & LaPolla, 1997; Van Valin, 2005, 2008) and the Lexical Constructional Model (LCM) (Mairal-Usón & Ruiz de Mendoza, 2008; Ruiz de Mendoza & Mairal-Usón, 2008; Ruiz de Mendoza, 2013; Ruiz de Mendoza & Galera, 2014). RRG is a current theory that accounts for an integrated description of grammar, meaning and function, and whose descriptive potential has been highlighted by scholars such as Kailuweit et al. (2018). The LCM is a construction grammar that presents a strong constructional layered typology that has been adopted by the knowledge base FunGramKB. Thus, in this section we will also outline the most relevant aspects for our research concerning the theoretical tenets of RRG and the LCM.

ARTEMIS (Automatically Representing Text Meaning via an Interlingua-based System) is an NLP prototype that has been implemented as a parsing device within the multipurpose[3] lexico-conceptual knowledge base FunGramKB and that can generate the syntactic and semantic representation of English sentences. FunGramKB stores conceptual, constructional, lexical and morphological information about the English language in independent but interrelated modules (see Figure 1):

3 The knowledge base FunGramKB is said to be multipurpose since it is multilingual in the sense that it can be used with several natural languages, and it can be implemented in different NLP tasks, such as machine translation, information retrieval and automatic summarising, dialogue-based applications such as question-answering, etc. (Periñán-Pascual & Arcas-Túnez, 2014).

Figure 1: The FunGramKB architecture (http://www.fungramkb.com/)



The required linguistic input that FunGramKB needs for the development of the lexical module (Lexicon) has been arranged in the knowledge base following the formal principles of the functional theory of RRG with regard to the lexical representation of predicates. Thus, in the Core Grammar component of the Lexicon in FunGramKB we can find the attributes that are used in the automatic building of the semantic and syntactic representation of sentences in ARTEMIS: *Aktionsart* ascription (verb class), macrorole assignment (Actor/Undergoer), status of variables, inventory of argumental constructions, etc. The development of the grammatical module in FunGramKB (Grammaticon) is mostly grounded on the constructional view of the LCM, whose layered structure of meaning construction (argumental level-1 constructions, implicational level-2 constructions, illocutionary level-3 constructions and discursive level-4 constructions) has allowed the integration of constructional meaning into RRG to deepen semantic processing (Periñán-Pascual, 2013a: 206). Similarly, the LCM notion of construction has been claimed to be more adequate for the computational requirements of the parser ARTEMIS (Periñán-Pascual, 2013a; Luzondo-Oyón & Ruiz de Mendoza, 2015; Fumero-Pérez & Díaz-Galán, 2017). Thus, the term “construction” is reserved for those constructions whose meaning is larger than the meaning of the building blocks conforming it and that are stored in the Grammaticon in FunGramKB. Structures whose meaning is fully compositional are, on the contrary, stored in the Lexicon of FunGramKB and should be referred to as “kernel constructs” as proposed by Luzondo-Oyón & Ruiz de Mendoza (2015).

ARTEMIS, as a syntactic and semantic parser, resorts to FunGramKB to obtain the lexical, semantic, syntactic and constructional information that is needed to transduce fragments of language into their conceptual logical structure (CLS) and syntactico-semantic representation, as represented in Table 1:

Table 1: The process of understanding natural language in ARTEMIS: phases and components (Periñán-Pascual & Arcas-Túnez, 2014)

<table>
<tr><th colspan="4">ARTEMIS ARCHITECTURE</th></tr>
<tr><th>PHASES</th><th colspan="3">COMPONENTS/MODULES</th></tr>
<tr><td>INPUT TEXT + PRE-PROCESSING</td><td colspan="3">Lemmatisation and tagging of word tokens</td></tr>
<tr><td rowspan="4">GRAMMAR BUILDING</td><td colspan="3">GRAMMAR DEVELOPMENT ENVIRONMENT (GDE):
2 THEORETICAL CONSTRUCTS</td></tr>
<tr><td colspan="3">1. Representation of feature-based structures as Attribute-Value Matrixes (AVMs) for grammatical units</td></tr>
<tr><td colspan="3">2. Feature-based production rules</td></tr>
<tr><td>Syntactic rules: they build the enhanced model of the Layered Structure of the Clause (LSC).</td><td>Constructional rules: they embed constructional schemata into the enhanced LSC (FunGramKB L1-Constructicon).</td><td>Lexical rules: they provide the morphosyntactic and semantic information of tokens (FunGramKB Lexicon and Ontology).</td></tr>
<tr><td>SYNTACTIC PARSING</td><td colspan="3">Generation of parse trees from the given input sentence</td></tr>
<tr><td>CONCEPTUAL LOGICAL STRUCTURE (CLS) EXTRACTION</td><td colspan="3">CLS CONSTRUCTOR
Enhanced text meaning representation of RRG logical structures</td></tr>
<tr><td>COREL (CONCEPTUAL REPRESENTATION LANGUAGE) SCHEME REPRESENTATION</td><td colspan="3">COREL SCHEME BUILDER
The CLS is modelled into a COREL scheme (formal language that formalises conceptual knowledge in FunGramKB).</td></tr>
</table>

We have used the example provided by Fumero-Pérez and Díaz-Galán (2017: 38-41), "Louise baked a cake for the kids", to illustrate the process that is followed in ARTEMIS in the *understanding* and transformation of natural language input into its equivalent grammatical and semantic structures. The first stage automatically separates the sentence components and assigns an Attribute-Value Matrix to each of them in which their semantic and morphosyntactic features are listed. Table 2 illustrates the output of this process for the word "baked":

Table 2: Attribute-Value Matrix of the word token "baked" (Fumero-Pérez & Díaz-Galán, 2017: 39)

INPUT TEXT	Louise had baked a cake for the kids.
GRAMMAR BUILDING	**1. Attribute-Value Matrix of the word token "baked"** Form baked Lemma bake POS verb Tense past Concept +BAKE_00

The second phase in the Grammar Building process involves the activation of production (syntactic, constructional and lexical) rules. Their task is, firstly, to "ascribe an appropriate syntactic/semantic contour to the sentences" (Fumero-Pérez & Díaz-Galán, 2017: 39) and, secondly, to generate a syntactic tree (syntactic parsing phase). Table 3.1 shows the feature-based syntactic rule that includes the semantic and syntactic information contained in the Core Grammar of the verb "bake". The *Aktionsart* of the predicate "bake" is active accomplishment (ACA) and the verb takes 2 variables, x (theme) and y (referent), that must fulfil the selection restrictions of the theme being +HUMAN_00 and the referent +FOOD_00:

Table 3.1: Feature-based syntactic rule for the predicate "bake" (Fumero-Pérez & Díaz-Galán, 2017: 39)

GRAMMAR BUILDING	Syntactic rule for "bake": [[Headword bake] [Core grammar [CLS [Aktionsart ACA] [Variables [Type x Role theme Macrorol Phrase Syntax Prefer +HUMAN_00] [Type y Role referent Macrorol Phrase Syntax Prefer +FOOD_00]]]]]

Table 3.2 illustrates the constructional rule for the verb "bake" when it participates in the For-Benefactive construction (FBEN).[4] The rule accounts for the fact the construction has added the constituent "for the kids", and this requires the inclusion of a new variable "w" with the role of Beneficiary and functioning as an argument adjunct. The construction has also modified the original aspectual value of "bake" (ACA), which is now a causative accomplishment (CACC). The causativity imposed by the FBEN construction triggers another change in the thematic role assigned to the variable "x", which is now an agent and not a theme (as shown in Table 3.1). Besides, the rule includes a realization constraint as regards the type of preposition that can introduce the prepositional phrase ("for") and a selection restriction,

4 The Core Grammar of the verb points to the different types of construction in which a verbal predicate can appear. In the case of the predicate "bake", apart from participating in the For-Benefactive construction, it can also appear in the "Unexpressed Second Argument Construction (The kitchen smelled so wonderful while they were baking); the Instrument Subject Construction (This oven bakes wonderful bread); the Material Subject Construction (This flour bakes a delicious loaf) and the Benefactive Object Construction (She baked them a cake)" (Fumero-Pérez & Díaz-Galán, 2017: 39).

+ORGANISM_00, that is defined in the FunGramKB ontology as “an animal, plant, human or any other living thing” (Fumero-Pérez & Díaz-Galán, 2017: 40).

Table 3.2: Feature-based constructional rule for the predicate “bake” in the For-Benefactive construction (Fumero-Pérez & Díaz-Galán, 2017: 40)

	Constructional rule for the For-Benefactive construction			
GRAMMAR BUILDING	L1-constr	[Type	FBEN]	
	CLS	[Aktionsart	CACC]	
		Variables	Type	x
			Role	agent
			[Type	y]
			Type	w
			Role	beneficiary
			Phrase	PP
			Syntax	argument adjunct
			Prep	for
			Prefer	+ORGANISM_00
	[COREL scheme		*(e1:<EVENT>) (f1:w) Beneficiary)]	

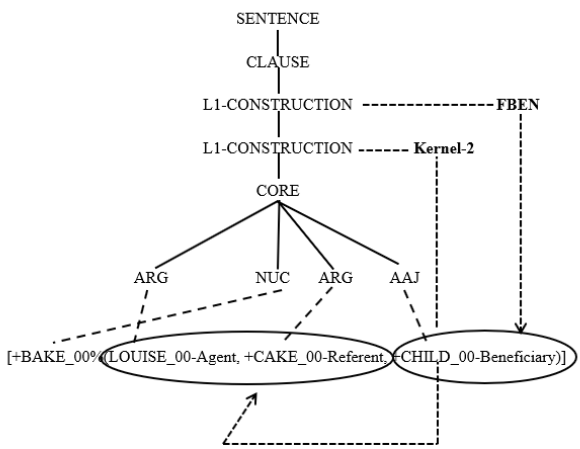
Finally, Table 3.3 presents the RRG lexical representation of the logical structure of the predicate “bake” in the example “Louise baked a cake for the kids”. As can be seen, the lexical representation accounts for the causative nature of the event that is introduced by the FBEN construction (Fumero-Pérez & Díaz-Galán, 2017: 40):

Table 3.3: Lexical representation for the predicate "bake" in the For-Benefactive construction (Fumero-Pérez & Díaz-Galán, 2017: 40)

GRAMMAR BUILDING	**Lexical representation:** RRG logical structure [[**do'**(Louise, ∅)] CAUSE [BECOME **bake'**(cake)]] PURP [BECOME **have'**(the kids, cake)]

All the information generated in the GDE is now used to yield the parsed tree for our example in the following phase of the parsing process (Table 4):

Table 4: Syntactic parsing phase of the example "Louise baked a cake for the child" (Fumero-Pérez & Díaz-Galán, 2017: 41)

SYNTACTIC PARSING	**Parsed tree: ARTEMIS syntax-semantic linkage** 

In the final phases of the parsing process (Table 5), the CLS is first extracted, and that same information is then presented in a purely semantic conceptual representation using the formal FunGramKB representation language COREL. The operators used in the CLS show that the illocutionary force of our example is "declarative", the tense is "past", the type of level-1 construction is the "FBEN" (which is a "kernel-2" construction that corresponds to monotransitive verbal predicates) and the *Aktionsart* is causative accomplishment “CACC”. At the end of the representation, the FunGramKB ontological concepts to which the different lexical items are linked are given:[5]

Table 5: The “CLS extraction” and “COREL scheme presentations” phases for the example “Louise baked a cake for the child” (Fumero-Pérez & Díaz-Galán, 2017: 41)

CLS Extraction	$<_{IF}{}^{DECL}<_{TENSE}{}^{PAST}<_{CONSTR\_L1}{}^{FBEN}<_{CONSTR\_L1}{}^{KER2}<_{AKT}{}^{CACC}$<[+BAKE_00(%LOUISE_00-Agent,+CAKE_00-Referent,+CHILD_00-Beneficiary)] >>>>>>[5]
COREL scheme representation	+(e1: +BAKE_00 (x1: %LOUISE_00)$_{THEME}$ (x2: +CAKE_00) $_{REFERENT}$ (f1: (e2: +DO_00 (x1)$_{AGENT}$ (e1)$_{REFERENT}$ (f2: +CHILD_00)Beneficiary)) Purpose)

In the implementation process of ARTEMIS, the RRG descriptive apparatus has had to undergo two important adaptations. On the one hand, in RRG, the representation of grammatical categories like aspect, negation and directionals (at the level of the nucleus), directionals, event quantification, modality and negation (at the level of the core) and status, tense evidentials and illocutionary force (at the level of the clause) is given in the Operator Projection that is represented in a distinct projection from the one representing predicates and their arguments (i.e., the Constituent Projection) (Van Valin, 2005: 12).[6] This Operator Projection has been substituted in ARTEMIS by Attribute-Value Matrixes (AVMs) (Cortés-Rodríguez & Mairal-Usón, 2016: 95) that capture the corresponding values for grammatical categories (see Table 6 for two examples), and that adhere to the principle of linearity of processing established by the computational parsing application “so that a tag or label can be assigned to each of the constituents in the sentence, and the machine can analyse them in a strict sequential order” (Martín-Díaz & González-Orta, 2020: 10). Additionally, the Operator Projection has also been replaced by unification mechanisms (Boas & Sag, 2012; Sag et al., 2003) where “morphosyntactic parsing is carried out jointly by a set of production rules and a number of feature unification operations intended to satisfy the structural and semantic constraints encoded in the AVMs” (Cortés-Rodríguez & Mairal-Usón, 2016: 96).

5 The symbol ‘%’ that can be found in the conceptual logical structures is used to introduce an ontological concept for specific entities in the conceptual model (Onomasticon) of FunGramKB (see Figure 1).

6 The Constituent Projection in RRG defines three syntactic units in the structure of the clause: the nucleus (which includes a verbal, an adjectival or a nominal predicate), the core (which contains the nucleus and its arguments), and the periphery (which includes constituents that are not predicate arguments).

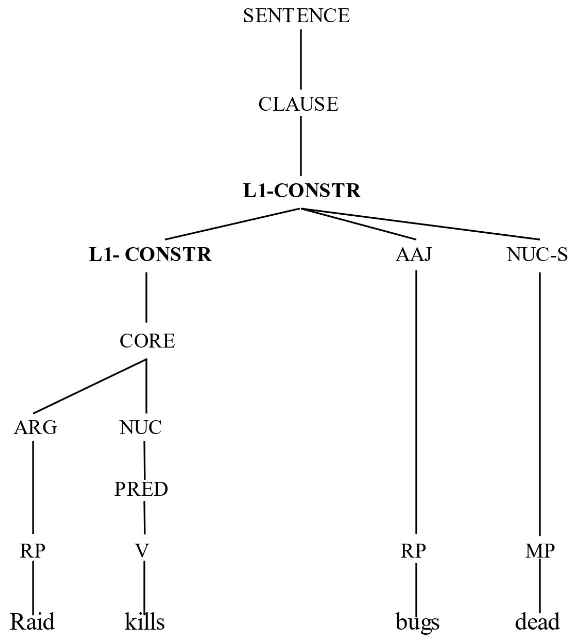
Table 6: Examples of the AVMs for the categories AUX and VERB (Cortés-Rodríguez & Mairal-Usón, 2016: 106-107)[7]

AVM for the category "auxiliary verb"	AVM for the category "verb"
<Category Type="AUX">	<Category Type="VERB">
<Attribute ID="Aspect" />	<Attribute ID="Aspect" />
<Attribute ID="Illoc" />	<Attribute ID="Concept" />
<Attribute ID="Num" />	<Attribute ID="Illoc" />
<Attribute ID="Per" />	<Attribute ID="Num" />
<Attribute ID="Tense"/>	<Attribute ID="Per" />
</Category>	<Attribute ID="Recip" />
	<Attribute ID="Reflex" />
	<Attribute ID="Template" />
	<Attribute ID="Tense"
	</Category>

On the other hand, and as put forward by Periñán-Pascual and Arcas-Túnez (2014), the RRG layered structure of the clause has been modified with the insertion of an intermediate constructional node, the level-1 construction node (L1-CONSTR), between the CORE and the CLAUSE nodes (see Figure 2 and the refined tree presented in Table 4), thus reflecting the constructional orientation of the LCM and their four-layered architecture of constructions, all of which are stored in the grammatical module (Grammaticon) of FunGramKB (see Figure 1).

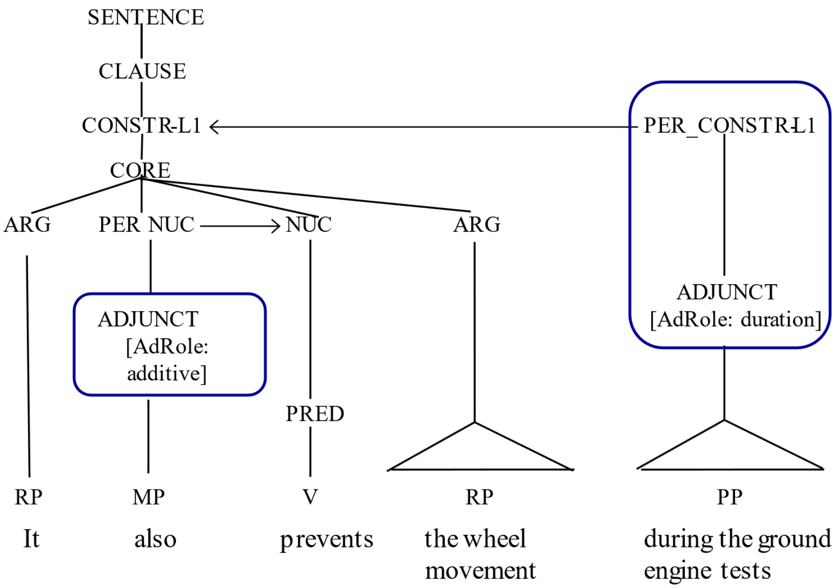
7 AVMs are encoded in XML format, similar to that of other platforms for the analysis of human language data, as is NLTK (Natural Language Toolkit; Bird et al., 2009).

Figure 2: The enhanced LSC of an English resultative level-1 construction



Adverbials in RRG are regarded as peripheral components that modify the three different levels of the Constituent Projection of the clause: in the nuclear periphery, focusing (*only, also, just*), degree (*completely, slightly*) and frequency adverbs (*daily, normally, never*) are located; in the core periphery, the following adverbial modifiers can be found: contingency (*although, in order to*), process (*because, apart from, loudly*), pace (*quickly*), space (*downstairs, above, from*) and temporal modifiers (*during, since, before*); finally, in the clause periphery we can place illocutionary (*briefly, frankly*), evidential (*apparently, presumably*) and epistemic adjuncts (*basically, certainly*) (see Rodríguez-Juárez & Cortés-Rodríguez (forthcoming) for a revised typology of adjuncts). As already mentioned, the addition of an intermediate constructional node (*L1-CONSTR*) in FunGramKB reflects the constructional orientation of the LCM, and since core adverbials share the same semantic typology and positional preferences as the adverbials found in level-1 argumental constructions, core adverbials will be reanalysed as peripheral constituents modifying the L1-CONSTR and, as a result, will be also referred to as CONSTR-L1 adjuncts or L1 adjuncts. Figure 3 shows an example of adverbials modifying the nuclear and L1-CONSTR nodes:

Figure 3: Tree of the enhanced LSC with the insertion of the CONSTR-L1 node of an Airbus sentence



After having summed up the relevant aspects of the grammatical models on which FunGramKB and ARTEMIS are grounded, and briefly outlined the architecture of the knowledge base and the parser, we will now move on to present the AVMs and the parsing rules that have had to be designed to respond to the conditions imposed by the ASD-STE100 controlled natural language that is used in the Airbus corpus.

3. Description of the corpus and methodology

The Airbus corpus[8] is made up of a collection of raw texts that belong to the domain of aircraft maintenance and the subdomain of aeronautical English–aircraft maintenance and that deal with instructions, descriptions and warning notices. It is a closed, synchronic and untagged corpus made up of 2,486 files / 6.697.387 bytes (xml format) that contains 687,345 word tokens (Felices-Lago & Alameda-Hernández, 2017: 109). The corpus has been written, as mentioned in the introduction of this paper, following the lexical and syntactic constraints established by the ASD-STE100 (2017) controlled language. Adverbs in ASD-STE100 are considered as a part of speech together with other seven parts of speech (verb, noun, pronoun, article, adjective, preposition and conjunction) and are briefly defined as "a word that modifies a verb, an adjective,

8 The Airbus corpus is at the disposal of our research group courtesy of Airbus in Seville.

or another adverb. It answers the questions, 'how?, 'where?', 'when', 'how often?' and 'how much?'" (ASD-STE100, 2021, Part 2 Dictionary: 2-0-4). Some orientation is also given as to how to form adverbs from adjectives: "Frequently (but not always), you can make an adverb from an adjective when you attach '-ly' ending to it. The comparative and superlative forms of adverbs are also made with 'more' and 'most'. Thus, they are not given in the dictionary" (ASD-STE100, 2021: 2-0-6). In fact, the words "more" and "most" are listed in the dictionary as independent approved words that can be used to form the comparative and superlative forms of adverbs. Other regular and irregular comparative and superlative adverbs formed by adding "-er" or "-est" are not approved in the ASD-STE100 dictionary. For example, the comparative adverb "later" is listed in the dictionary, but in lower case letters, which indicates that it is not an approved word that should be substituted by the adverb "subsequently", as in "Make sure that the tool is *subsequently* available (and not "available *later*") for the installation procedure" (ASD-STE100, 2021, Part 2 Dictionary: 2-1-L2). ASD-STE100 (2021) includes 182 approved examples of adverbs, prepositions and conjunctions (Appendix 1) that can be codified as adjuncts realised by adverb phrases ("The MLG door uplock closes *mechanically* and opens *hydraulically*"), prepositional phrases ("Make sure that there are no objects or persons *below the aircraft*") and adverbial clauses ("Put the sensor (6) in the correct position *while you keep the spacer (10) and sensor (2) in position*"), respectively.

Prior to the design of the rules for the internal realization of adjuncts and for their placement in the layered structure of the clause (syntactic rules), as well as to the creation of the catalogue of Attribute-Value Matrixes, a preliminary study was carried out devoted to the study of the semantic and syntactic analysis of a representative sample of adjuncts in the Airbus corpus (Rodríguez-Juárez & Cortés-Rodríguez, forthcoming). To be precise, a sample of 99 adjuncts (67 examples of adverbs, 24 prepositions that could be used to introduce prepositional phrases, and 8 conjunctions introducing adverbial clauses) were analysed in 7,603 example sentences according to the position they occupied in the sentence and to their semantic input.[9] Of these 99 adjuncts (see Appendix 2), 84.84% were examples of approved words included in the Dictionary of ASD-AST100, and only 15.15% were examples of adverbs that are not approved in this STE (already, completely, at the same time, even, except, far, in case of, in general, individually, inside, never, normally, partially, remotely, sometimes), which shows that, although the corpus mostly adheres to the rules of the specification document ASD-STE100 (2017), some deviations can be attested.

9 For a detailed account of the methodology followed in the compilation of the corpus of Airbus adjuncts, see Rodríguez-Juárez & Cortés-Rodríguez (forthcoming).

The design of the parsing rules will reflect the positional and peripheral preferences of the two adjunct types (nuclear adjuncts (NUC) and core/L1-CONSTR adjuncts (CORE/L1)) that were registered in the Airbus corpus and that are presented in Table 7. The possible adjunct positions within the sentence are distributed along four main positions: the extra-clausal positions (i.e., the left-detached position (LDP) and the right-detached position (RDP,) as in "*In this condition* (LDP), the EBCU is in standby mode, *until the crew member applies pressure to one of the brake pedals* (RDP)) and the initial ("*Always* use gloves for protection") , medial ("It *also* adjusts the hydraulic flow in the return operation") and final ("Lower the trailing arm (25) positions *carefully*") positions.

Table 7: Positions and peripheral preferences of adverbials in the Airbus corpus (Rodríguez-Juárez & Cortés-Rodríguez, forthcoming)

PERIPHERAL PREFERENCES	POSITIONS				
	LDP	INITIAL	MEDIAL	FINAL	RDP
+High (100%-51%)				CORE/L1 (58.12%)	
± Middle (50%-20%)	NUC (24.59%)		NUC (48.23%) CORE/L1 (24.79%)	NUC (25.53%)	
–Low (19%-0%)	CORE/L1 (15.26%)	NUC (0.95%) CORE/L1 (0.90%)			NUC (0.71%) CORE/L1 (0.93%)

Thus, the previous research into the semantics and syntax of adverbials in the Airbus corpus served as a basis for the design of the rules for the internal realization of adjuncts in the Airbus corpus and for their placement in the layered structure of the clause (syntactic rules) as well as for the elaboration of the catalogue of Attribute-Value Matrixes that are needed in the implementation of this CNL in the parser ARTEMIS. The integration of these production rules in the GDE of ARTEMIS will ultimately be helpful for an effective and more detailed automatic parsing of sentences in which optional peripheral constituents appear, thus increasing the explanatory potential of this prototype.

4. Our contribution to the implementation of Airbus adjuncts in ARTEMIS

One of the main aims of this study is to contribute to the development of the NLP prototype ARTEMIS by providing, on the one hand, the catalogue of Attribute-Value Matrixes (AVMs) for peripheral constituents (Section 4.1), and, on the other hand, the syntactic rules for both the internal realization of the different peripheries (Section 4.2) and for the localization of the peripheries at the different layers of the constituent structure that are necessary for the parsing of the syntactic and semantic representation of Airbus adjuncts (Section 4.3). It is important to bear in mind that, differently from other syntactic parsers, the rules that we have designed are based on a solid linguistic model and are aimed at retrieving not only the syntactic but also the semantic structure of a given fragment of language.

4.1. Catalogue of AVMs

As already mentioned in Section 2, for a satisfactory and effective formalization of adverbials in ARTEMIS, a methodological adaptation of the RRG descriptive apparatus had to be embraced that consisted in the replacement of the RRG Operator Projection by AVMs, which are feature-bearing structures that encode the selectional and semantic information of the different types of grammatical constituents in the format of attributes and values. The AVMs for the category Adjunct and the sentence nodes for the nuclear peripheries (PER_NUC) and the core/level-1peripheries (PER_CORE/L1) have been refined with the addition of the attribute “weight” in the case of nuclear and core peripheries:

ADJUNCT

```
<Category Type=”ADJUNCT”>
  <Attribute ID=”AdjunctRole” />
  <Attribute ID=”concept” />
  </Category>
```

PER_NUC

```
<Category Type=”PER_NUC”>
  <Attribute ID=”AdjunctRole” />
  <Attribute ID=”concept” />
  <Attribute ID=”weight” />
  </Category>
```

PER_CORE/L1

<Category Type="PER_L1">

<Attribute ID="AdjunctRole" />

<Attribute ID="concept" />

<Attribute ID="weight" />

</Category>

A closer look at the attributes included in these AVMs shows that the values under the attribute "Adjunct Role" display all the possible semantic types that have been distinguished for the three types of adverbials in the different layers of the LSC (see Rodríguez-Juárez & Cortés-Rodríguez, forthcoming), as can be seen below, where the type of periphery has been facilitated in brackets ((N) for nuclear, (C) for core and (CL) for clausal peripheries):

<Attribute ID="**AdjunctRole**" obl="+" num="s">

<Value>**Additive**</Value> (N)

<Value>**Beneficiary**</Value> (C)

<Value>**Company**</Value> (C)

<Value>**Concession**</Value> (C)

<Value>**Conditional**</Value> (C)

<Value>**Definite frequency**</Value> (N)

<Value>**Degree/Amplifiers**</Value> (N)

<Value>**Degree/Diminisher**</Value> (N)

<Value>**Direction**</Value> (C)

<Value>**Distance**</Value> (C)

<Value>**Duration**</Value> (C)

<Value>**Epistemic**</Value> (CL)

<Value>**Evidential**</Value> (CL)

<Value>**Exception**</Value> (C)

<Value>**Illocutionary**</Value> (CL)

<Value>**Indefinite freq.**</Value> (N)

<Value>**Instrument**</Value> (C)

<Value>**Limiter**</Value> (N)

<Value>**Location/Position**</Value> (C)

<Value>**Manner**</Value> (C)

<Value>**Means**</Value> (C)

<Value>**Pace**</Value> (C)

<Value>**Path**</Value> (C)

<Value>**Purpose**</Value> (C)

<Value>**Reason**</Value> (C)

<Value>**Result**</Value> (C)

<Value>**Source**</Value> (C)

<Value>**Span**</Value> (C)

<Value>**Time position**</Value> (C)

</Attribute>

The value for the attribute "concept" in the AVMs indicates that the parser has to address the FunGramKB core ontology to find and retrieve the concept associated with the lexical entry of the head word functioning as adjunct; this is encoded as an instruction by means of the string [FIND: core > concept > concept | CHECK: %\w*]. The attribute "concept" is then expressed as follows:

<Attribute ID="**Concept**" obl="*" num="s"

<Value>[FIND: core > concept > concept | CHECK: %\w*]</Value>

</Attribute>

Finally, we have added an attribute, "weight", with 6 values, that is going to determine the likelihood or probability of occurrence of an adverbial in a particular position of the LSC. In the AVM, the weight values are open, but they are codified at the very beginning of the syntactic parsing where each position is given a specific value based on the analysis of the Airbus positional preferences and frequencies that we have established for the different peripheral constituents (Rodríguez-Juárez & Cortés-Rodríguez, forthcoming). The weight that is attached to each position is dependent on the frequency of occurrence of adjuncts in that particular position. Thus, the higher the frequency of occurrence, the less marked the assignment of an adjunct to that position is, and, as a result, the lower the weight that is assigned to this position in the ARTEMIS parsing rules:

```
<Attribute ID="Weight" obl="*" num="1">
<Value>0</Value>
<Value>1</Value>
<Value>2</Value>
<Value>3</Value>
<Value>4</Value>
<Value>5</Value>
<Value>6</Value>
</Attribute>
```

We reproduce here (Table 8) the weight-based priority that has been assigned to each position and for each type of adjuncts as a result of the different rates of frequency that have been registered in the previous study of the Airbus corpus (Rodríguez-Juárez & Cortés-Rodríguez, forthcoming):

Table 8: Assignment of weight to the different positions per adjunct type based on frequency rates in the Airbus corpus (Rodríguez-Juárez & Cortés-Rodríguez, forthcoming)

POSITION	Core/L1 adjuncts	WEIGHT	NUC adjuncts	WEIGHT
RDP	0.93%	6	0.71%	6
FIN	58.12%	3	25.53%	5
MED	24.79%	5	48.23%	4
INI	0.90%	6	0.95%	6
LDP	15.26%	6	24.59%	5

WEIGHTS	FREQUENCIES
1	83.34% - 100%
2	66.68% - 83.33%
3	50.02% - 66.67%
4	33.36% - 50.01%
5	16.70% - 33.35%
6	0.00% - 16.69%

The assignment of a weight-based approach to positional preferences is substantiated by psycholinguistic evidence supporting the use of frequency factors by the human sentence parser in order to iron out local ambiguity (cf. Pickering & van Gompel, 2006). Additionally, Periñán-Pascual and Arcas-Túnez (2014: 186) suggest that "it would be more effective to apply the 'weight-based priority' from the beginning of the syntactic parsing with the purpose of minimizing global syntactic ambiguity". The inclusion of the weight attribute for adverbials guarantees that it will be activated at the earlier stages in parsing operations.

4.2. Parsing rules for the internal realization of airbus peripheries in ARTEMIS

The rules for the internal realization of the different peripheries incorporate and distribute the values assigned to the different attributes of the AVMs (Table 9). In the case of adjuncts, the syntactic rule includes the attributes "concept" and "adjunct role" with all the possible semantic types. The rule also codifies its different realizations as modifier phrases (MP) (like *also*), prepositional phrases (PP) (like *behind the handle*), referential phrases (RP) (like *today*) and clauses (CL) (like *until the hinge pin is installed*):

Table 9: Rule for the internal realization of adjunct peripheries in ARTEMIS

ADJUNCT[**adjunctrole**=additive|amplifier|beneficiary|company|concession|condition al| definitefrequency|diminisher|direction|distance|duration|exception|indefinitefreque ncy|instrument|limiter|location|manner|means| pace|path|purpose|reason|result|sou rce|span| timeposition, **concept**=?]

→ **MP** || **PP**[prep=?p, concept=?] || **RP**[cnt=?, concept=?, def=?, dei=?, num=?, per=?, quant=?] || **CL** [akt=?, concept=?, emph= ?, illoc= ?, sta=?, t=?, tpl=?]

The rule for the internal codification of nuclear peripheries shows a restriction in the array of possible semantic types admitted in this periphery and incorporates the attribute "weight" with all the six possible values (Table 10). The rule also indicates that nuclear peripheries are realised by adjuncts:

Table 10: Rule for the internal realization of nuclear peripheries in ARTEMIS

PER_NUC[**adjunctrole**=additive| degreeamplifier|definitefrequency|degreediminisher| indefinitefrequency| limiter, **concept**=?, **weight**=1|2|3|4|5|6]
→ **ADJUNCT** [adjunctrole=additive|definitefrequency|degree|indefinitefrequency|limit er, concept=?, weight=1|2|3|4|5|6]

The rule for core/L1 peripheries also includes a restricted and distinct list of semantic types that can occur at this level and shows that this periphery may be realised by one, two or even by the concatenation of three adjuncts (Table 11). The rule also specifies the attribute "concept" and the potential "weight" that can be assigned to each position:

Table 11: Rule for the internal realization of core/L1 (concatenated) peripheries in ARTEMIS

1. ONE ADJUNCT

PER_L1[adjunctrole=beneficiary|company|concession|conditonal|direction|distance|duration|exception|instrument|location|manner|means|pace|path|purpose|-reason|result|source|span| timeposition, **concept**= ?, **weight**=1|2|3|4|5|6] →

ADJUNCT[adjunctrole=beneficiary|company|concession|conditonal|direction|distance|duration|exception|instrument|location|manner|means|pace|path|purpose-|reason|result|source|span| timeposition, concept: ?, weight=1|2|3|4|5|6] ||

2. CONCATENATION OF TWO ADJUNCTS

PER_L1[adjunctrole=beneficiary|company|concession|conditonal|direction|distance|duration|exception|instrument|location|manner|means|pace|path|purpose|-reason|result|source|span| timeposition, **concept**= ?, **weight**=1|2|3|4|5|6] →
ADJUNCT[adjunctrole=beneficiary|company|concession|conditonal|direction|distance|duration|exception|instrument|location|manner|means|pace|path|purpose-|reason|result|source|span| timeposition, concept: ?, weight=1|2|3|4|5|6]
ADJUNCT[adjunctrole=beneficiary|company|concession|conditonal|direction|distance|duration|exception|instrument|location|manner|means|pace|path|purpose-|reason|result|source|span| timeposition, concept: ?, weight=1|2|3|4|5|6]

3. CONCATENATION OF THREE ADJUNCTS

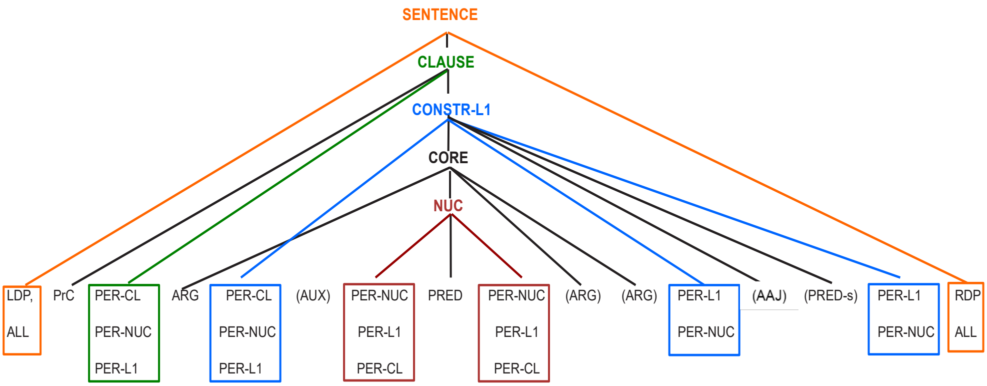
PER_L1[adjunctrole=beneficiary|company|concession|conditonal|direction|distance|duration|exception|instrument|location|manner|means|pace|path|purpose|-reason|result|source|span| timeposition, **concept**= ?, **weight**=1|2|3|4|5|6] →

ADJUNCT[adjunctrole=beneficiary|company|concession|conditonal|direction|distance|duration|exception|instrument|location|manner|means|pace|path|purpose-|reason|result|source|span| timeposition, concept: ?, weight=1|2|3|4|5|6]
ADJUNCT[adjunctrole=beneficiary|company|concession|conditonal|direction|distance|duration|exception|instrument|location|manner|means|pace|path|purpose-|reason|result|source|span| timeposition, concept: ?, weight=1|2|3|4|5|6]
ADJUNCT[adjunctrole=beneficiary|company|concession|conditonal|direction|distance|duration|exception|instrument|location|manner|means|pace|path|purpose-|reason|result|source|span| timeposition, concept: ?, weight=1|2|3|4|5|6]

4.3. Parsing rules for the localization of Airbus peripheries in the layers of the LSC

This section deals with the process followed in the elaboration of the syntactic rules for the localization of the peripheries in the different layers of the constituent structure of the clause that will be stored in the GDE of the parser ARTEMIS. The peripheries are represented as daughter constituents that belong to the central layers in the LSC. In this way, the computational requirements for the parsing of these constituents in the syntactic rules are met since these rules can only capture linear ordering and immediate dominance (Cortés-Rodríguez & Rodríguez-Juárez, 2019: 74). The scheme provided in Figure 4 for standard English shows that the detached extra-clausal positions (LDP/RDP) are attached to the Sentence, the initial positions are assigned to the Clause and also to the L1-Constructional layer, the medial positions are assigned to the Nucleus, and the final positions are assigned to the L1-Constructional layer. These dominance relations have to be reflected in the parsing rules:

Figure 4: Peripheries and positional preferences in the abstract LSC for standard English (with dominance relations) (Cortés-Rodríguez & Rodríguez-Juárez, 2019: 75)



Detached position	Initial position: initial & end variants	Medial position: medial & end variants	Final position: initial & final variants	Detached position

Before presenting the rules for the localization of adjuncts in the different layers of the LSC, it should be reminded that in the controlled natural language of the Airbus corpus, there is only one instance of a clausal adjunct (*possibly*) in the sentence "The aircraft is possibly not parallel to the ground", as opposed to standard English. This is again one difference in comparison with standard English where clausal subjective adjuncts like *frankly*, *presumably* or *cleverly* are quite frequent. Since ASD-STE100 is a

CNL that aims at providing information as objectively as possible, these types of clausal subjective adjuncts do not occur. Thus, our rules will only display the localization of core/L1 and nuclear peripheral constituents.

At the layer of the nucleus, the parsing rule for Airbus nuclear adjuncts is formulated as follows:

NUC → (PER_L1) $_{\text{Weight:5}}$ (PER_NUC) $_{\text{Weight:4}}$ **PRED** (PER_NUC) $_{\text{Weight:4}}$ (PER_L1) $_{\text{Weight:5}}$

The parentheses indicate optionality of constituents since otherwise the number of possibilities would multiply their linearization possibilities. This rule shows that in the Airbus corpus at the layer of the NUC we can locate nuclear and core/L1 adjuncts in medial positions (as shown in Figure 4), each with a particular weight. As mentioned above, in the AVMs, the weight values are left open, but in the parsing rules each position is given a specific weight. So, in the case of L1 adjuncts in medial position, the weight assigned is 5, which shows that this occurrence is less frequent (24.79%) than the one registered for nuclear adjuncts with a weight of 4 (48.23%). Examples extracted from the Airbus corpus of the possible combinations are given below:

(PER_NUC) PRED:

7. Each system *continuously* monitors the other system.

(PER_L1) PRED:

8. This test sequence *automatically* does a check of the actuators.

PRED (PER_L1):

9. Connect *mechanically* the brake unit to the main wheel.

PRED (PER_NUC):

10. Let the landing gear extend *fully*.

PRED (PER_NUC) (PER_L1):

11. Their routing is *also near to the structure of the leg assembly* ...

At the layer of the core/L1 periphery, the adjuncts can be located both in initial and in final positions, with different arrangements as can be seen in the rule:

PER_L1/CORE → (ARG) (PER_NUC) $_{\text{Weight:6}}$ (PER_L1) $_{\text{Weight:6}}$ (AUX) **PRED** (ARG) (ARG) (PER_L1) $_{\text{Weight:3}}$ (PER_NUC) $_{\text{Weight:5}}$ (AAJ) (PRED-s) (PER_L1) $_{\text{Weight:3}}$ (PER_NUC) $_{\text{Weight:5}}$

Below we show examples of a nuclear periphery in initial position after the argument and before the auxiliary (example 12), of a nuclear periphery in final position (example 13) and of a concatenation of a nuclear and a core/L1 adjunct in final position after the argument (example 14):

(ARG) (PER_NUC) (AUX) PRED:

12. The hitch pin *only* can be installed if the kneeling lock valve is correctly closed.

PRED (ARG) (PER_NUC):

13. The Normal Braking System operates the brake units *usually*.

PRED (ARG) (PER_NUC) (PER_L1):

14. Do not push the pin (11) *completely through the lug*.

At the layer of the clause, the nuclear and core/L1 peripheries have to be located in initial positions:

CL → (PER-NUC) $_{\text{Weight:6}}$ (PER_L1) $_{\text{Weight:6}}$ **CONSTR-L1**

We have registered examples of nuclear and core/L1 adjuncts in this position but, as can be seen from the weight assigned to them, they are not found very often (weight 6). No examples of a concatenation of two distinct peripheries before the L1-construction have been registered:

(PER_NUC) CONSTR-L1:

15. *Always* the top bleed valve (8) of the related brake unit (1) must be used ...

(PER_L1) CONSTR-L1:

16. *When the aircraft is on-ground* the MLG shock absorber is compressed.

At the level of the sentence, examples of all the possible arrangements have been registered in the Airbus corpus:

SENTENCE → (LDP) CLAUSE (RDP)

In the left and right detached positions, which are pragmatically motivated positions that are usually separated by commas or pauses from the rest of the clause, we have found examples of both nuclear and L1 adjuncts:

LDP → (PER_L1) $_{\text{Weight:6}}$ (PER_NUC) $_{\text{Weight:5}}$

(PER_L1):

17. *In normal mode*, the system sends all the fault messages to ...

(PER_NUC):

18. *In general,* these materials are flammable, poisonous and ...

RDP → (PER_L1) $_{\text{Weight:6}}$ (PER_NUC) $_{\text{Weight:6}}$

(PER_L1):

19. The ADCN interchanges data of the Normal Braking System (*through the BACS software).*

(PER_NUC):

20. This type of equipment (...) can cause damage to equipment, *specially to: electrical equipment ...*

We have also registered examples of both LDP and RDP together in the same sentence:

(LDP) CLAUSE (RDP):

21. *In this condition,* the EBCU is in standby mode, *until the crew member applies pressure to one of the brake pedal assemblies.*

Despite the introduction of weighted options of realization of peripheral types in a given position, syntactic ambiguity will still persist, which will be translated in more than one final parse tree for a given adverbial. However, each tree will be provided with a different weighted value corresponding to its probability to become the winning option. In those cases of cooccurrence of several adverbials in a same sentence, the sum of the weighted values will determine which is the most probable overall syntactic analysis. This option enables the parser to minimise global syntactic ambiguity by solving all cases of local ambiguity in the structure.

5. Conclusions

Controlled natural languages present differences with respect to natural languages in the sense that they try to be simpler by avoiding ambiguity and implicit meaning and, as a result, aim at the comprehensibility and readability of technical texts. These simplified languages make them suitable candidates to be used in the first phases of the computational implementation of natural language processing. In the present study, we have chosen the ASD-STE100 CNL used in the Airbus corpus and have focused on the behaviour of Airbus adjuncts to test the validity of the parser ARTEMIS. Thus, our main aim to enrich the research conducted so far into the development of the grammatical module (GDE) of ARTEMIS through the design of the production rules for the different constituents of the clause has been fulfilled with our proposal of the set of AVMs and parsing rules for ASD-STE100 adverbial units, which have had to be adapted to respond to the reality imposed by the Airbus controlled natural language. In order to account for the large flexibility and variety of adjuncts in terms of the place they can occupy in the sentence and their semantic variability, we have incorporated a weight factor that will facilitate the automatic processing and computational parsing of Airbus adjuncts, thus achieving our secondary aim.

With this research we have contributed to completing the studies that have already been carried out on the development of the GDE in ARTEMIS and on the design of the AVMs and the syntactic rules for ASD-STE100 phrasal and clausal constituents, in which no analysis of adjuncts had been offered so far. Although our findings are limited in the sense that they are restricted to the adverbs of the ASD-STE100 CNL, we believe that they can serve as a stepping-stone towards the analysis of these and other constituents in other English-based CNLs that are relevant to the field of computational linguistics. On the other hand, English-based CNLs are in fact subsets of their "base language" (English) (Khun, 2014: 123), and, consequently, the work done so far can also be valid to do further research into the eventual generation of the logical conceptual structures of natural language expressions when we move from controlled natural languages into non-controlled natural ones. In this process, the scope of research should be widened by the analysis of data relative to the position and behaviour of clausal subjective adverbials (e.g. *frankly*, *possibly*, etc.) in standard English, as they are not used in texts produced in ASD-STE100.

6. References

Aerospace and Defence Industries Association of Europe (2017). *ASD-STE100 Simplified Technical English. Specification ASD-STE-100. International specification for the preparation of technical documentation in a controlled language*. Issue 7, January 2017. Brussels.

Aerospace and Defence Industries Association of Europe (2021). *ASD-STE100 Simplified Technical English. Specification ADE-STE-100. International specification for the preparation of technical documentation in a controlled language.* Issue 8, April 30, 2021. Brussels. https://asd-ste100.org/

Bird, S., Klein, E & Loper, E. (2009). *Natural Language Processing with Python*. Sebastopol: O'Reilly Media Inc.

Boas, H.C., & Sag, I. (eds.) (2012). *Sign-Based Construction Grammar.* Stanford: CSLI.

Clark, P., Harrison, P., Murray, W.R. & Thompson, J. (2010). Naturalness vs. predictability: a key debate in controlled languages. *Proceedings 2009 Workshop on Controlled Natural Languages (CNL'09).*

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

Cortés-Rodríguez, F.J. & Rodríguez-Juárez, C. (2018). Parsing phrasal constituents in ASD-STE100 with ARTEMIS. *Voprosy Kognitivnoy Lingvistiki*, *3*, 97-109.

Cortés Rodríguez, F.J. & Rodríguez-Juárez, C. (2019). The syntactic parsing of ASD-STE100 adverbials in ARTEMIS. *Revista de Lingüística y Lenguas Aplicadas*, *14*, 59-79.

Díaz-Galán, A. (2018). Deep parsing for the aviation industry: Adjusting ARTEMIS for parsing simple clauses in ASD-STE100. *Voprosy Kognitivnoy Lingvistiki*, *3*, 83-96.

Díaz-Galán, A. & Fumero-Pérez, M.C. (2016). Developing parsing rules within ARTEMIS: The case of *do* auxiliary insertion. In C. Periñán-Pascual & E.M. 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-Galán, A. & Fumero-Pérez, M.C. (2020). An account of constructions in ASD-STE100: Formalizing non-propositional meaning in aviation instructional texts. *Revista Electrónica de Lingüística Aplicada, 19*, 24-41.

Felices-Lago, A. & Alameda-Hernández, A. (2017). The process of building the upper-level hierarchy for the aircraft structure ontology to be integrated in FunGramKB. *Revista de Lenguas para Fines Específicos, 23*(2), 86-110.

Fuchs, N.E., Kaljurand, K. & Kuhn, T. (2008). Attemp to Controlled English for knowledge representation. *Reasoning Web, LNCS*, 5224, 104-124.

Fumero-Pérez, M.C. & 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.

Fumero-Pérez, M.C. & Díaz-Galán, A. (2019). Designing the lexical rules for the parsing of ASD-STE100 function words in ARTEMIS from a Role and Reference Grammar perspective. *Journal of English Studies*, *17*, 149-174.

González-Orta, M. & Martín-Díaz, A. (2022). ARTEMIS: Parsing non-peripheral complex sentences in ASD-STE100. *Onomázein*, *56*, 80-99.

Kailuweit, R., Künkel, L. & Staudinger, E. (eds) (2018). *Applying and Expanding Role and Reference Grammar.* Freiburg: Freiburg University Library [NIHIN Studies].

Kuhn, T. (2014). A survey and classification of controlled natural languages. *Computational Linguistics*, *40*(1), 121-170.

Luzondo-Oyón, A. & Ruiz de Mendoza Ibáñez, F.J. (2015). Argument structure constructions in a natural language processing environment. *Language Sciences*, *48*, 70-89.

Mairal-Usón, R. & Ruiz de Mendoza Ibáñez, F.J. (2008). New challenges for lexical representation within the lexical-constructional model (LCM). *Revista Canaria de Estudios Ingleses*, *57*, 137-155.

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

Martín-Díaz, M.A. (2018). Parsing *wh*-interrogative sentences within ARTEMIS. *Revista Electrónica de Lingüística Aplicada*, *17*(1), 78-98.

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

Martín-Díaz, M.A. & González-Orta, M. (2020). Clausal arguments and peripheries in ASD-STE100: The parsing of subordination in ARTEMIS. *Revista Electrónica de Lingüística Aplicada*, *19*(1), 1-23.

Periñán-Pascual, C. (2012). En defensa del procesamiento del lenguaje natural fundamentado en la lingüística teórica. *Onomázein, 26*, 13-48.

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

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

Periñán-Pascual, C. & Arcas-Túnez, F. (2014). The implementation of the FunGramKB CLS Constructor. In B. Nolan & C. Periñán-Pascual (eds.), *Language*

Processing and Grammars: The Role of Functionally Oriented Computational Models (pp. 165-196). Amsterdam: John Benjamins,

Pickering, M.J. & van Gompel, R.P.G. (2006). Syntactic parsing. In M.J. Traxler & M.A. Gernsbacher (eds.), *Handbook of Psycholinguistics* (pp. 455-503). Amsterdam: Elsevier.

Rodríguez-Juárez, C.& Cortés-Rodríguez, F.J. (forthcoming, 2022). A Role and Reference Grammar account of adjuncts in the Airbus Corpus: A quantitative-based study. *Atlantis*, *44* (2).

Ruiz de Mendoza Ibáñez, F.J. (2013). Meaning construction, meaning interpretation and formal expression in the Lexical Constructional Model. In B. Nolan & E. Diedrichsen (eds.), *Linking constructions into functional linguistics: The role of constructions in grammar* (pp. 231-70). Amsterdam: John Benjamins.

Ruiz de Mendoza Ibáñez, F.J. & Mairal-Usón, R. (2008). Levels of description and constraining of factors in meaning construction: an introduction to the Lexical Constructional Model. *Folia lingüística: Acta Societatis Linguisticae Europaeae*, *42*(2), 355-400.

Ruiz de Mendoza Ibáñez, F.J. & Galera Masegosa, A. (2014). *Cognitive Modeling. A Linguistic Perspective*. Amsterdam: John Benjamins.

Sag, I.A., Wasow, T. & Bender, E.M. (2003). *Syntactic Theory: Formal Introduction*. Stanford: CSLI Publications.

Schwitter, R. (2010). Controlled natural languages for knowledge representation. *Colin 2010: Poster Volume*, 1113-1121.

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

Van Valin, R.D. (ed.) (2008). *Investigations of the Syntax-Semantics-Pragmatics Interface*. Amsterdam: John Benjamins.

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

White, C. & Schwitter, R. (2009). An update on PENG Light. *Proceedings of ALTA*, 80-88.

APPENDICES

Appendix 1

List of approved adverbs, prepositions, and conjunctions in ASD-STE100 (2017):

ADVERBS

1. accidentally
2. accurately
3. across
4. aft
5. again
6. almost
7. also
8. always
9. apart
10. approximately
11. automatically
12. axially
13. back
14. brightly
15. carefully
16. chemically
17. clearly
18. clockwise
19. constantly
20. continuously
21. correctly
22. counterclockwise
23. diagonally
24. differently
25. digitally
26. dimly
27. directly
28. down
29. downstream
30. easily
31. electrically
32. electromagnetically
33. electronically
34. equally
35. externally
36. first
37. forward
38. freely
39. frequently
40. fully
41. gradually
42. here
43. horizontally
44. how
45. hydraulically
46. immediately
47. in
48. inboard
49. incorrectly
50. independently
51. initially
52. internally
53. intermittently
54. irregularly
55. in one (TN = technical noun) … and then the other
56. in progress
57. last
58. laterally
59. lightly
60. linearly
61. locally
62. longitudinally
63. loosely
64. magnetically
65. manually
66. mechanically
67. moderately
68. momentarily
69. more
70. much
71. no
72. not
73. off
74. on
75. only
76. out
77. outboard

78. outdoors
79. overboard
80. permanently
81. pneumatically
82. possibly
83. quickly
84. radially
85. randomly
86. rearward
87. regularly
88. safely
89. same
90. satisfactorily
91. slowly
92. smoothly
93. specially
94. structurally
95. subsequently
96. suddenly
97. sufficiently
98. symmetrically
99. temporarily
100. then
101. there
102. thus
103. through
104. tightly
105. together
106. too
107. unsatisfactorily
108. unusually
109. up
110. upstream
111. vertically
112. very
113. visually
114. yes

PREPOSITIONS

1. abaft
2. about
3. above
4. across
5. adjacent to
6. after
7. aft of
8. against
9. along
10. around
11. as
12. at
13. away from
14. because of
15. before
16. behind
17. below
18. between
19. by
20. down
21. downstream of
22. during
23. for
24. forward to
25. from
26. in
27. inboard of
28. in front of
29. into
30. minus
31. near
32. of
33. off
34. on
35. onto
36. opposite
37. outboard of
38. out of
39. over
40. plus
41. through
42. thru
43. to
44. until
45. up
46. upstream of
47. with
48. without

CONJUNCTIONS

1. after
2. although
3. and
4. as
5. as … as
6. because
7. before
8. but
9. if
10. or
11. since
12. than
13. that
14. until
15. unless
16. when
17. where
18. while

APPENDIX 2

List of word types (99) that can be realised as adverbial expressions (codified as adverbs, prepositional phrases or clauses); raw representative sample (7,603); final sample (5,180) after not valid instances were eliminated (manual filtering process) (Rodríguez-Juárez & Cortés-Rodríguez, forthcoming).

AIRBUS CORPUS		REPRES. SAMPLE	FINAL SAMPLE
WORD TYPES	WORD TOKENS	WORD TOKENS	WORD TOKENS
99	**67,556**	**7,603**	**5,180**
above	17	16	6
accurately	5	5	5
across	3	3	3
after	141	104	82
again	92	75	75
against	16	15	13
along	25	24	21
already	2	2	1
also	172	120	120
although	1	1	1
always	45	40	40
away	96	77	11
apart (from)	1	1	1
approximately	130	98	2
around	39	36	11
as	105	83	61

AIRBUS CORPUS		REPRES. SAMPLE	FINAL SAMPLE
WORD TYPES	WORD TOKENS	WORD TOKENS	WORD TOKENS
99	**67,556**	**7,603**	**5,180**
at the same time	134	100	100
automatically	45	40	40
because	36	33	33
before	830	270	258
behind	3	3	1
below	144	106	46
between	420	205	92
by	97	78	19
carefully	70	60	60
clearly	42	38	38
clockwise	25	24	23
completely	2	2	2
continuously	6	6	6
correctly	483	219	219
counterclockwise	20	19	19
directly	152	110	110
down	520	226	12
downstream	10	10	4
during	614	242	242
easily	16	15	15
electrically	24	23	22
equally	1	1	1
even	3	3	1
except	5	5	5
externally	1	1	1
forward	268	160	1
freely	6	6	6
far	8	8	2
for	5908	375	375

AIRBUS CORPUS		REPRES. SAMPLE	FINAL SAMPLE
WORD TYPES	WORD TOKENS	WORD TOKENS	WORD TOKENS
99	**67,556**	**7,603**	**5,180**
from	3438	358	90
fully	197	132	37
hydraulically	27	25	20
if	6945	378	335
immediately	102	81	75
in	9880	384	384
in case	2	2	2
in general	83	69	83
incorrectly	3	3	3
independently	8	8	7
individually	2	2	2
initially	1	1	1
inside	26	24	6
internally	2	2	2
lightly	5	5	1
linearly	1	1	1
manually	30	28	28
mechanically	30	28	28
momentarily	4	4	4
much	87	71	71
near	700	255	246
never	1	1	1
normally	1	1	1
off	621	243	29
on	15780	390	156
only	230	146	7
onto	3	3	3
out	76	64	37
over	18	17	3

AIRBUS CORPUS		REPRES. SAMPLE	FINAL SAMPLE
WORD TYPES	WORD TOKENS	WORD TOKENS	WORD TOKENS
99	**67,556**	**7,603**	**5,180**
partially	7	7	7
permanently	4	4	4
possibly	1	1	1
rearward	2	2	1
remotely	1	1	1
safely	5	5	5
since	1	1	1
slowly	64	55	55
smoothly	2	2	2
some times	5	5	5
specially	9	9	1
structurally	3	3	2
suddenly	9	9	9
sufficiently	2	2	2
temporarily	8	8	8
then	619	243	2
through	485	219	207
x times	34	31	26
to	13929	389	389
until	135	101	101
usually	5	5	5
when	1849	329	299
while	48	43	43
with	1231	302	131
without	12	12	4

APPENDIX 3

List of abbreviations:

AAJ	Argument Adjunct
ACA	Active Accomplishment
ACE	Attempto Controlled English
AdRole	Adjunct Role
ARG	Argument
ASD-STE100	AeroSpace and Defence-Simplified Technical English Specification
AUX	Auxiliary
AVM	Attribute-Value Matrix
CACC	Causative Accomplishment
CL	Clause
CLN	Controlled Natural Language
CLS	Conceptual Logical Structure
CORE/L1	Core/Level-1 CONSTRUCTION Adjuncts
COREL	Conceptual Representation Language
CPL	Computer Processable Language
FBEN	For-Benefactive construction
FIN	Final position
GDE	Grammar Development Environment
Iloc	Illocutionary Force
INI	Initial position
L1	Level 1
L1-CONSTR	Level-1 Construction
LCM	Lexical Constructional Model
LDP	Left-Detached Position
LSC	Layered Structure of the Clause
MED	Medial position
MP	Modifier Phrase

NLP	Natural Language Processing
NUC	Nucleus / Nuclear Adjunct
NUC-S	Secondary Nucleus
Num	Number
obl	Obligatory
PENG	Processable English
Per	Person
PER-CL	Clausal Peripheries
PER_L1/ CORE	Level-1/Core Peripheries
PER_NUC	Nuclear Peripheries
PP	Prepositional Phrase
PRED	Predicate
RDP	Right-Detached Position
Recip	Recipient
Reflex	Reflexive
RP	Referential Phrase
RRG	Role and reference Grammar
STE	Simplified Technical English
V	Verb