ATLANTIS
Journal of the Spanish Association of Anglo-American Studies
44.2 (December 2022): 20-44
e-ISSN 1989-6840
DOI: http://doi.org/10.28914/Atlantis-2022-44.2.02

A Role and Reference Grammar Account of Adjuncts in the Airbus Corpus: A Quantitative-Based Study

Carolina Rodríguez-Juárez
Universidad de Las Palmas de Gran Canaria
carolina.rodriguez@ulpgc.es

Francisco J. Cortés-Rodríguez
Universidad de La Laguna
fcortes@ull.edu.es

This paper presents the results of a quantitative study of adjuncts in the Airbus corpus carried out within the theoretical framework of Role and Reference Grammar (RRG). We describe the positional behaviour of these peripheral constituents in the Layered Structure of the Clause and postulate scales of positional and peripheral preferences, based on frequency distribution, in the Airbus controlled natural language (CNL). The results obtained were compared with a previous study on adjunct preferences and positions in Natural English to check for changes in these scales due to the nature of the texts written in this CNL. We also aim to contribute to the development of the RRG analysis of adverbials by offering a detailed semantic typology and a description of the syntax of these peripheral constituents grounded in empirical and quantitatively based data that will serve as a basis for the parsing of adverbials in the computational processing of CNLs.

Keywords: adjuncts; Airbus corpus; controlled natural languages; corpus linguistics; peripheries; Role and Reference Grammar

. . .

Estudio cuantitativo sobre los adjuntos en el corpus Airbus basado en la Gramática del Papel y la Referencia

Este trabajo ofrece los resultados de un estudio cuantitativo sobre los adjuntos en el corpus Airbus realizado dentro del marco de la Gramática del Papel y la Referencia (GPR) en el que describimos el comportamiento posicional de los constituyentes periféricos en la estructura estratificada de la cláusula y postulamos escalas de preferencias posicionales y de tipos de constituyentes periféricos, basadas en distribución de frecuencias, en el lenguaje natural controlado (LNC) del corpus Airbus. Los resultados obtenidos se comparan con los de un estudio previo sobre preferencias y posiciones de los adjuntos en inglés estándar, con el fin de comprobar si puede haber cambios en estas escalas debido a la naturaleza de los textos escritos en un LNC. También nos proponemos contribuir al desarrollo del análisis de los adverbios en la GPR ofreciendo una tipología semántica y una detallada descripción de la sintaxis de estos constituyentes periféricos sustentada en datos empíricos y cuantitativos que servirán como base para el parseado de los constituyentes adverbiales en el procesamiento computacional de los LNCs.

Palabras clave: adjuntos; corpus Airbus; lenguaje natural controlado; lingüística de corpus; periferias; Gramática del Papel y la Referencia

1. Introduction

The study of the syntactic and semantic behaviour of adverbs in English has been widely addressed in various descriptive grammatical models that provide semantic typologies, classifications in terms of their formal realisation and grammatical functions as well as positional norms that describe the preferred order of these constituents in the clause (Quirk et al. 1985; Downing and Locke 1992; Hudleston and Pullum 2002, to name a few). Functional grammars like Dik's Functional Grammar (henceforth FG; Dik 1997a, 1997b) and Van Valin's Role and Reference Grammar (henceforth RRG; Van Valin and LaPolla 1997; Van Valin 2005) also propose a formal representation of how the different types of adverbs modify the different layers of the underlying structure of the clause in present-day English.

Not so much has been investigated, however, about the behaviour of adverbial constituents in English-based controlled natural languages (CNLs). A CNL is an umbrella term that covers the large number of constructed languages that have been designed for use in specific environments and disciplines and for different purposes. These constrained natural languages should use "a well-defined subset of a language's grammar and lexicon" together with "the terminology needed in a technical domain" (Kittredge 2003, qtd. in Khun 2014, 122), but they do not necessarily have to exhibit exactly the same characteristics, since some are inherently ambiguous whereas others are very precise; some resemble natural languages, while others are closer to programming languages (Kuhn 2014, 121) and others are constructed in such a way that they "can be accurately and efficiently processed by a computer" (Fuchs and Schwitter, qtd. in Khun 2014, 122). Most CNLs are based on English (Spaggiari et al. 2005, 107) and their importance has been widely attested in the field of computational linguistics (Kuhn 2014, 122). CNLs pursue readability, comprehensibility and translatability (Spaggiari et al. 2005, 107) and, as a consequence, are more restrictive than natural languages. Therefore, in these controlled languages we should expect some limitations and restrictions in the syntactic and semantic behaviour of the different constituents.

This research deals with the study of adverbials in a CNL and has three main goals. In the first place, we aim to present the results of a corpus-based study on the semantic and syntactic behaviour of adverbial constituents in the Airbus corpus, an aircraft maintenance manual that includes procedural and descriptive texts and safety instructions which has been written in Simplified Technical English (STE) in accordance with the instructions and restrictions in syntax and vocabulary that are compiled in the specification document ASD-STE100 (2017). As expected, the language used in this corpus is simpler than Natural English and we aim to explore whether significant differences can be perceived between the use of adverbials in Natural English and in the CNL used in the Airbus ASD-STE100 corpus.

Our second goal aims at developing the basic proposal of analysis of adverbials presented in the theoretical framework of RRG (Van Valin and LaPolla 1997; Van Valin 2005), and, specifically, at providing a detailed analysis of the semantics and syntax of

adjuncts in the layered structure of the clause (LSC) in accordance with Van Valin's reanalysis of such constituents (2005, 19-21), taking into account that, although this reanalysis adds relevant information, it still only provides a skeletal framework, which needs to be expanded to provide a complete overview of the semantic and syntactic description of peripheral constituents.

Thirdly, we aim to provide empirical and quantitatively based criteria for the effective parsing of adverbials in the computational processing of ASD-STE100, as is being implemented by some natural language processing (NLP) prototypes like ARTEMIS (Automatically Representing Text Meaning via an Interlingua-based System). This is a parsing device designed within the multiple lexico-conceptual knowledge base FunGramKB (Periñán-Pascual 2012, 2013; Periñán-Pascual and Arcas-Túnez 2014) to obtain the syntactic and semantic representation of linguistic structures (Periñán-Pascual and Arcas-Túnez 2014). The parser is currently being applied to the CNL ASD-STE100 due to its simple and unambiguous syntax, and for this same reason it is used in this preliminary validation stage as "a test bench to achieve the eventual parsing of natural English" (Fumero-Pérez and Díaz-Galán 2019, 152). With respect to the computational processing of Airbus adverbials, it is crucial to consider that the positional mobility and the semantic variability of this type of constituent make it difficult to accurately process them automatically.

This corpus-based study is part of a longitudinal research project whose initial findings were presented in Cortés-Rodríguez and Rodríguez-Juárez (2019). In this phase of the research project, we will expand on the previous study of adjuncts, which was based on evidence from Natural English and from a brief study of adverbials in a small sample of 213 examples taken from the Airbus corpus, to consider the whole corpus of Airbus ASD-STE100 and conduct a complete quantitative analysis of adverbials in the LSC so that we can fulfil the following specific aims:

1. To describe the positional behaviour of peripheral constituents in the Airbus CNL.
2. To test the scales of peripheral preferences and positions for Natural English proposed in Cortés-Rodríguez and Rodríguez-Juárez (2019) on the CNL of the Airbus corpus, in order to check whether Airbus STE involves changes on these scales as a result of the nature of the texts written in this CNL.
3. To propose an indicator of the relative weight of probability of occurrence of the different types of adjuncts in a given position, with the aim of facilitating the automatic processing of these constituents in any NLP task.

The remainder of this article is structured as follows. Section two gives a general account of adjuncts in English in terms of their semantic content and positional preferences and formulates a proposal of an enhanced semantic and syntactic typology of adjuncts. Section three briefly outlines the grammatical aspects of RRG that are necessary in this investigation. Section four deals with the description of the corpus

and the methodology followed in the corpus-based study. Section five reports the key findings of the quantitative analysis. In section six we offer some concluding remarks.

2. An Account of Adjuncts in Natural English: A Proposal for an Enhanced Adjunct Typology

Numerous descriptive grammars have addressed the semantics and syntax of English adverbials, postulating different classifications that reflect their semantic input and syntactic behaviour. As regards their semantic content, Quirk et al. (1985, 479-85) outline seven main semantic roles—space, time, process, respect, contingency, modality and degree—, all but one of which are divided into subtypes. Additionally, they classify adverbials into four categories in terms of their grammatical function: adjuncts and subjuncts, which "are relatively integrated with the structure of the clause," versus disjuncts and conjuncts, which "have a more peripheral relation in the sentence" (Quirk et al. 1985, 440). Downing and Locke also categorise adverbials into adjuncts, disjuncts and conjuncts (1992, 58-64) and present six broad types of meaning with their respective subtypes: circumstantial, process, modal, degree, focusing and conjunctive (551-52). Huddleston and Pullum (2002) classify adjuncts into two groups that show the orientation of the adjunct with respect to the verb phrase or the clause and which include various semantic types. Thus, on the one hand, verb phrase-oriented adjuncts include the semantic classes of manner, means, instrument, act-related, spatial location, source, goal, path, direction, extent, degree, temporal location, duration, aspectuality, frequency, serial order, purpose, reason, result, concession and condition (576, 670-724); while on the other hand, clause-oriented adjuncts are realised by the semantic types of domain: modality, evaluation, speech-act related and connective (576-80, 765-79).

Functional grammars have also addressed the role of adverbials in natural languages in terms of their semantic input. In the theory of FG, Dik (1997a, 1997b) refers to *adverbials* as lexical satellites that modify the different levels of the underlying clause structure according to the type of further optional information they attach to the state of affairs (1997a, 87): a) predicate satellites—modifying the core predication; b) predication satellites—modifying the extended predication; c) attitudinal—modifying the proposition; and d) illocutionary satellites—modifying the clause. As a result, he presents an exhaustive list of semantic types distributed throughout the different layers (1997a, 64-67, 225-32, 234-36, 243-45, 297-99, 304-307).

Even though there is a more recent variant of this model (Functional Discourse Grammar, Hengeveld and Mackenzie 2008), the grammatical component remains very similar to FG, especially with respect to clausal constituents. By way of example, Hengeveld et al. (2019, 271) list the following types of semantic domains for operators and modifiers at the interpersonal and representational levels of the grammatical component, where the semantic and pragmatic aspects of linguistic expressions are encoded:

1. Interpersonal level: Communicative status of the move (e.g., *in sum*); communicative status of the act (e.g., *in addition*); stylistic properties of the act (e.g., *briefly*); illocutionary manner (e.g., *frankly*); subjective attitude (e.g., *fortunately*); reportativity (e.g., *reportedly*).
2. Representational level: Propositional attitude (e.g., *possibly*); evidence (e.g., *apparently*); order of episodes (e.g., *first*); absolute time (e.g., *yesterday*); relative time (e.g., *after that*); reality status (e.g., *hardly*); event quantification (e.g., *twice*); manner (e.g., *beautifully*); aspect (e.g., *continuously*).

In RRG, these peripheral elements are known as adjuncts and are conceived as non-arguments that can be codified as: a) adverbs—non-phrasal adjuncts; b) prepositional phrases—phrasal adjuncts (Van Valin 2005, 19); or c) adjunct clauses that can either be introduced by predicative prepositions that take clausal objects or be marked by conjunctions such as *because*, *if* or *although* (Van Valin 2005, 194). Adjuncts act as modifiers of the different layers in the constituent projection of the LSC: nucleus, core and clause (Van Valin 2005, 19-21, 183). In section three, the LSC will be introduced in more detail, but it is important to highlight now that in RRG the constituents of the clause are: a) the nucleus—realised by a verbal, adjectival or nominal predicate; b) the core—predicate + arguments; and c) the clause—core + periphery (non-obligatory) adjuncts (Van Valin 2005, 4). Van Valin (2005, 19) provides a reduced list of semantic types per layer: a) nucleus adjuncts like aspectual adverbs (*continuously*); b) core adjuncts such as temporal adverbs (*yesterday*); c) pace adverbs (*quickly*) and manner adverbs (*carefully*); and d) clause adjuncts like epistemic adverbs (*probably*) and evidential adverbs (*evidently*).

In these descriptive and functional grammars, restrictions on the position of these peripheral constituents are outlined and linked to their semantics, grammatical role and formal realisation in the sentence. Quirk et al. (1985, 491-501) distinguish: a) the *initial* position—preceding any other clause element; b) the *medial* position—between the subject and the verb, with two subvariants: 1) *initial medial* position—between the subject and the operator; and 2) *end medial* position—before the main verb in a verb phrase; and c) the *end* position—after all obligatory elements with an end variant, i.e. *initial end* position—when "end-focus […] makes preferable an obligatory element in clause-final position despite the presence of an adverbial" (Quirk et al. 1985, 499). Grounded in a corpus-based study, they show how positional norms vary according to the formal realisation of the adverbial element, there being a tendency for adverbials to appear mostly in end position: prepositional phrases (79%), closed-class items (42%), finite clauses (66%) and noun phrases (74%). The medial position is mostly occupied by open-class adverbs (42%), whereas initial positions are more frequently occupied by finite clauses (31%) and closed-class items (25%; Quirk et al. 1995, 489-90, 500-501).

Huddleston and Pullum (2002, 575) also differentiate three linear positions of adjuncts in clause structure, namely: a) front position—before the subject; b) end position—after the verb; and c) central position—between the subject and the

verb in clauses headed by a lexical verb or by an auxiliary verb, although in the latter case it is common to find the adverb after the verb, which leads to problems distinguishing it from end position. Although the placement of adjuncts can be conditioned by variation in use and features of context, style, prosody, etc., some positional generalisations can be drawn (Huddleston and Pullum 2002, 576). Thus, clause-oriented adjuncts are "less closely associated with the VP constituents and less likely to be positioned in the VP or adjacent to the VP," whereas VP-oriented adjuncts are "more closely associated with the VP constituents, and more likely to be positioned in the VP or adjacent to the VP" (Huddleston and Pullum 2002, 576). There are even cases of adverbials that can have both types of orientation. For example, the adjunct *frankly* can occupy the front position (*Frankly*, *I'm just not interested*) or be placed after the catenative auxiliary when it is a speech act-related adverb and has a clause orientation (*I would frankly want a lot more money than that for it*), but it can also occupy the central position when it follows the catenative auxiliary and is a manner adverb with a verb-phrase orientation (*I would frankly explain to him what the position was*; Huddleston and Pullum 2002, 575-76).

Dik's FG comments sparingly on the position of satellites in the predication. Basically, he mentions that the unmarked position of manner adverbs is after the verb, i.e., they prefer "a non-initial position" (1997a, 226), whereas level three attitudinal satellites (*probably*) prefer an initial position and are "marked off from the rest [of the predication] by prosodic inflections" (1997a, 226, 298).

In RRG, the position of peripheral constituents in the different layers of the clause seems to be constrained by their interaction with operators as well as by the level they modify.[1] Thus, for instance, in the case of manner adverbs, Van Valin (2005, 20) states that these adverbs "interact in an important way with the tense operator" and if they are placed before it they will be "construed as clausal modifiers" (as in *Ruth cleverly hid the cash*), whereas if they are placed after it, they cannot (as in *Ruth hid the cash cleverly*; Van Valin 2005, 20). In instances where there is a concatenation of adverbs, their position is restricted by the LSC in the sense that "adverbs related to more outer operators occur outside of adverbs related to more inner operators" (Van Valin 2005, 20). Van Valin notes that the position of adverbs in the sentence is not fixed, as it is in the case of operators, but does point out some scope restraints that "require that the nuclear adverb be closer to the verb than the core adverb, and likewise for the core adverb with respect to the clausal adverb" (2005, 21). In the case of prepositional phrases, they normally occur after the core and, in those instances in which they are placed before it, they must be located in the pre-core slot or left-detached position (Van Valin 2005, 21). Adjunct adverbial clauses—prepositions

[1] Dik (1997a, 51-52) also conceives the parallelism between operators and satellites in the building up of the structure of the clause in the sense that they both modify the state of affairs by adding the same type of semantic content at each layer, albeit expressed in different ways: by grammatical and lexical means, respectively.

that take clauses as objects—also occur in the periphery of the core, whereas adverbial clauses—introduced by conjunctions—must occur in the periphery of the clause (Van Valin 2005, 194).

This brief review of the treatment of adverbs in some well-established descriptive grammars and the two functional models mentioned, RRG and FG, have provided us with a framework to articulate a more fine-grained typology of adjuncts in RRG, which has been built following two types of criteria: 1) the position they occupy in the RRG layered structure of the clause; and 2) their semantic input. Thus, adjuncts are first assigned to one of the layers of the LSC that they can modify on the basis of their most frequent and natural positional tendencies, after which they are classified within each layer into different types and subtypes depending on the semantic information that they convey. Table 1 shows the revised typology of adjuncts, enriched with a more detailed classification of adjunct subtypes and with more examples:[2]

Table 1. A layered clause-oriented and semantic-based typology of adjuncts in English

CLAUSE MODIFIERS		
Adjunct Types		**Examples**
Illocutionary		approximately, briefly, broadly, frankly, honestly, literally
Evidential		apparently, evidently, in my experience, obviously, presumably, reportedly, supposedly
Epistemic		actually, basically, certainly, clearly, definitely, for sure, of course, probably, really, undoubtedly
CORE MODIFIERS		
Adjunct Types	**Subtypes**	**Examples**
Contingency	Concession	although X, despite X, in spite of X, though X
	Conditional	If X
	Purpose	for X, in order to-inf/in order that, so as to X, to-inf.
	Reason	because, because of X, for X, on account of X
	Result	so (that) X

[2] See Díaz-Jorge (2017) and Cortés-Rodríguez and Rodríguez-Juárez (2019) for an initial though less detailed outline of this classification.

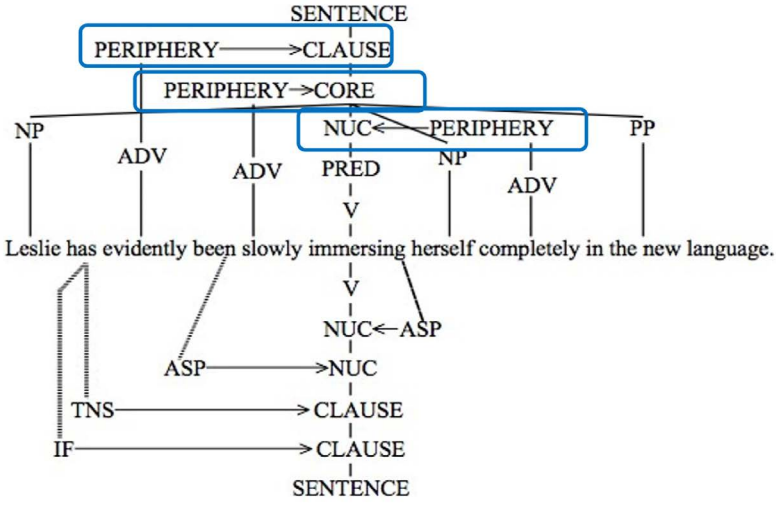
Process	Beneficiary	for X
	Company	with X, accompanied by X
	Exception	apart from, X, except X
	Manner	loudly, recklessly, silently
	Means/Instrument	by X, through X, with X
Pace		quickly, slowly
Space	Direction	down X, downwards, downstairs, left/right, to X, towards X
	Distance	a long way, X miles/kilometres, for a few miles, the whole distance
	Location/Position	above X, downstairs, in X, here, near X, on X, over X, there, where X, with (+ animate)
	Path	across X, along X, past X, through X
	Source	from X, out of the X
Temporal	Duration	during X, for X, lately, recently
	Span	by X, by the time X, since X, until/till X, up to X
	Time position	after X, again, at night, before X, from X to Y, in June, in the morning, on Monday, then ('at that time'), today, tonight, tomorrow, when/while X

NUCLEUS MODIFIERS		
Adjunct Types	**Subtypes**	**Examples**
Focusing	Restrictives/Limiters	alone, already, approximately, exactly, just, merely, only, simply
	Additives	also, apart from X, as well, even, further, too
Degree	Amplifiers	completely, fully, (very) much, totally, still
	Diminishers	a bit/little, almost, barely, hardly, slightly
Frequency	Definite frequency	annually, daily, hourly, monthly, nightly, once, twice, weekly, X times
	Indefinite 1 (High frequency)	frequently, often, regularly, repeatedly
	Indefinite 1 (Usual/Mid frequency)	commonly, generally, normally, on most days, sometimes, usually
	Indefinite 2 (Low frequency)	hardly ever, rarely, occasionally, scarcely, seldom
	Indefinite 2 (Zero frequency)	never, nevermore, no longer
	Indefinite 3 (Continual/ Universal frequency)	always, constantly, continually, incessantly, permanently

3. A Brief Overview of Role and Reference Grammar (RRG)

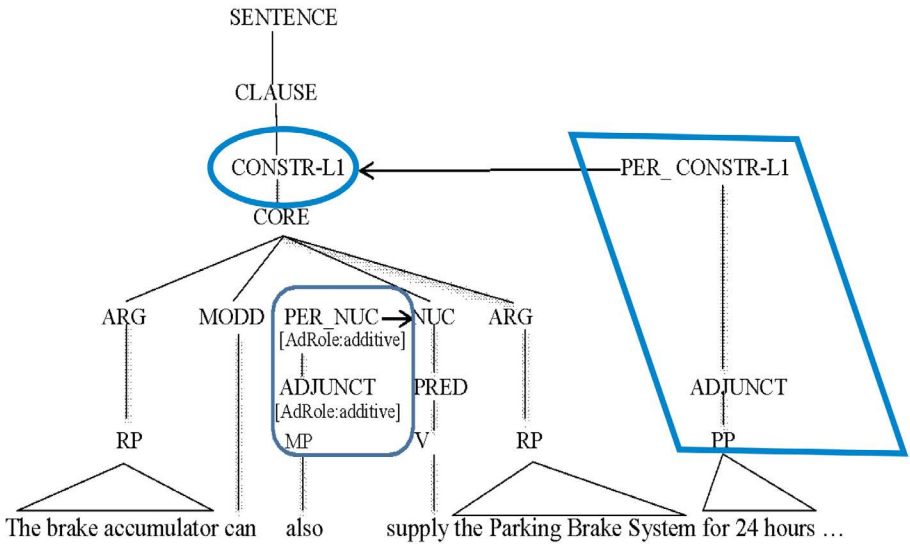
In this study on the description of adjuncts in the Airbus corpus framed within the theory of RRG, we need to pay special attention to their conception of the LSC, which depicts how semantics is projected onto syntax and comprises the constituents that make up a clause. Thus, the core contains the predicate (i.e., the nucleus, which can be verbal, adjectival or nominal) and its arguments—obligatory constituents—, while the periphery contains non-arguments of the predicate—optional constituents—, i.e., the so-called adjuncts. The core and the periphery components constitute the clause (Van Valin 2005, 4-5). In this research, we aim specifically to analyse the semantics and syntax of peripheral constituents in the LSC following Van Valin's reanalysis (2005) of such constituents, where adverbials are no longer conceived as acting as modifiers of a single layer (the CORE node, Van Valin and LaPolla 1997) but are distributed throughout the three different layers in the constituent projection of the LSC (Van Valin 2005, 19). In fact, although Van Valin and LaPolla (1997) mention in passing that adverbial constituents can be found in the three layers of the clause (168-69), this fact is not reflected in any of the analyses that are provided in the volume. It is not until Van Valin (2005) that actual instances of analyses of sentences illustrating peripheries modifying the different layers of the clause can be found for the English language (22, 195, 232). As a result of this reanalysis, the syntactic analysis must contemplate a nuclear periphery, a core periphery and a clause periphery, as shown in figure 1, where the three peripheral modifiers have been inserted in boxes.

Figure 1. The Layered Structure of the Clause (LSC) in English, including peripheral elements (Van Valin 2005, 22)



In our analysis, the RRG descriptive apparatus shows a modification in the layered structure of the clause where an intermediate constructional node—CONSTR-L1—between the CORE and the CLAUSE nodes (see figure 2) has been inserted. This modification reflects the constructional orientation of the Lexical Constructional Model (LCM) proposed by Ruiz de Mendoza (2013) and Ruiz de Mendoza and Galera (2014) and their four-layered architecture of constructions (level one, or L1, argumental constructions; L2, implicational constructions; L3, illocutionary constructions; and L4, discursive constructions), which has been followed in the design of the knowledge base FunGramKB.[3] In the case of adverbials, and as supported by Cortés-Rodríguez and Rodríguez-Juárez (2019, 67), "it seems feasible to reanalyse core adverbials as peripheral units modifying the CONSTR-L1 layer" since their semantic typology and their positional preferences are almost identical in both cases. Thus, in this study, core adverbials will also be referred to as CONSTR-L1 adjuncts, or simply as L1 adjuncts.

Figure 2. Refined tree of the LSC with the insertion of the CONSTR-L1 node (Cortés-Rodríguez and Rodríguez-Juárez 2019, 69)



This modification was necessary to adapt the RRG descriptive apparatus to the computational requirements of ARTEMIS (Periñán-Pascual and Arcas-Túnez 2014), as was a second adaptation related to the operator projection, which represents the different

[3] The knowledge base FunGramKB has been designed following the principles of the functional theory of RRG (Van Valin and LaPolla 1997; Van Valin 2005) and the LCM (Ruiz de Mendoza 2013; Ruiz de Mendoza and Galera 2014) and stores conceptual, constructional, lexical and morphological information about the English language.

grammatical categories at the different levels of the clause. This operator projection has been substituted in ARTEMIS by "feature-bearing matrixes and unification mechanisms" (Cortés-Rodríguez and Rodríguez-Juárez 2019, 60), the so-called Attribute Value Matrixes (AVMs) that lodge the corresponding values for grammatical categories and comply with the principle of linearity of processing established by the computational parsing application (Martín-Díaz and González-Orta 2020, 10).

4. Description of the Corpus and Methodology

The Airbus corpus is a closed, synchronic and untagged corpus made up of 2,486 files (xml format) that contain 687,345 tokens in total and includes a collection of raw written texts that belong to the domain of aircraft maintenance and the subdomain of aeronautical English. These texts have been written following the lexical and syntactic constraints established by the ASD-STE controlled language and deal with instructions (how to use… and safety procedures), descriptions (elements, technical data and systems) and warning notices (Felices-Lago and Alameda-Hernández 2017, 109).

The methodology followed in this quantitative study was built in various phases. The first stage—corpus gathering and selection of the representative sample in the Airbus corpus—consisted in manually selecting all the examples of adverbs (non-phrasal adjuncts), prepositions and conjunctions that could potentially form part of prepositional phrasal adjuncts or adverbial clauses from the list of 2,742 word types that constitute the population of the complete Airbus corpus. The extraction tool used for this purpose was AntConc (Anthony 2019), a freeware corpus analysis toolkit for concordancing and text analysis that generates the Word List of all the words (word types) in the corpus presented in an alphabetical list together with the number of occurrences per word (word tokens). At this very first stage, some manual filtering processes were carried out across the whole population so that the sample would only include relevant examples of adjuncts. For example, word types such as *piano*, which can be both adjectives and adverbs, were not included, since, in the three instantiations of this word in the corpus, *piano* worked as an adjective; the same happened with *outboard* and *straight*. This manual selection process gave us the population of word types that could be realised as adverbial expressions in the Airbus corpus: ninety-nine word types and 67,556 word tokens.

The second phase in the selection of the sample consisted in calculating the size that the sample of examples should theoretically have by applying statistical methods that calculate the number of examples that can be considered as homogeneous and representative of the whole population (Blecua et al. 1999, 63). Thus, it was necessary to calculate the minimum number of examples that would be required for each word type for the sampling results to be generalisable to the whole population with a margin of error of 0.05 (5%) and a 95.5% confidence level (Neuber 1980, 48-49). Table 2 shows the equation used to calculate the sample size of the different subgroups (García-Ferrando 1985, 142).

Table 2. Equation for the calculation of the size of the representative sample

$n = \frac{z^2 \cdot N \cdot p \cdot q}{N \cdot E^2 + z^2 \cdot p \cdot q}$	[n = sample size; N = population size; z^2 = confidence level [z = 2 (95.5%)]; E = margin of error (0.05); p = individual probability of the phenomenon (0.5); q = complementary probability (0.5)]

The different values that the equation takes in each of the specific calculations for the ninety-nine adverbial word tokens determined the theoretical sample size (n) = 7,603 word tokens. Once we started with the analysis of the sample, we still had to discard some examples that were not instances of adjuncts modifying verbal predicates or clauses but were in fact examples of post-noun and pre-noun modification, adverbs modifying adjectives and other adverbs, prepositions or adverbs working as particles in phrasal verbs or even incomplete sentences. After this manual filtering process, we were left with the real and accurate sample to be analysed, which comprised ninety-nine word types occurring in 5,180 example sentences.

In the third phase of the study, we extracted the number of concordances that were representative of each adverbial word type using the AntConc Concordance Tool, which "shows search results in a 'KWIC' (KeyWord in context format)" (Anthony 2019, n.p.). In order to select the representative number of sentences that had to be analysed for each word type (after applying the formula stated in table 2), we extracted the examples in the same order as they are given in the list of concordances in AntConc. For example, in the case of *above*, there are seventeen instantiations in the whole corpus, but the representative sample is sixteen, which means that we had to analyse the first sixteen concordances provided. Of those sixteen examples, ten were not valid for our analysis because *above* functioned as a postmodifier of a noun phrase (*in the step above; with a dashed line above it*), so only six examples were analysed in the end, as explained in the second phase of our methodology.

The fourth stage consisted in conducting the individual analysis of the adjuncts in each of the example sentences. For each adjunct, the semantic type it belonged to (manner, degree, frequency, etc.) was noted in an Excel document, as well as the level of the clause to which it was assigned (clausal, core/L1, nuclear) and the position (left detached position (LDP) or right detached position (RDP), initial, medial, final) that it occupied in the sentence so that we could register the frequency of occurrence in order to postulate scales of positional and peripheral preferences for the adverbs of the Airbus corpus.

In the last phase of the study, we established a comparison between the results obtained in the analysis of adjuncts in the complete Airbus corpus and the results obtained in the study of the reduced sample of the Airbus corpus (213 instances) and of Natural English (Cortés-Rodríguez and Rodríguez-Juárez 2019).

5. Quantitative Analysis of Airbus Adjuncts

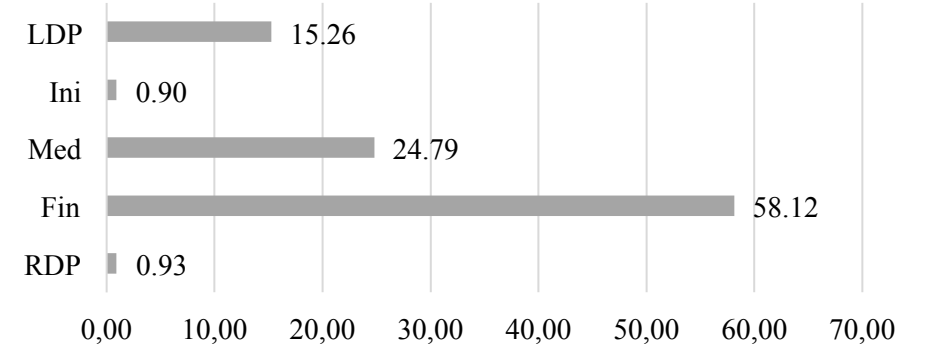
5.1. Positional Preference Depending on Type of Peripheral Constituent in the Airbus Corpus

In this section we present the results for the positional preferences of the different types of peripheries (clausal, core/L1, nuclear adjuncts). As regards clausal peripheries, only one example of an epistemic clausal adjunct in medial position, *possibly*, was found in the whole corpus (*The aircraft is possibly not parallel to the ground*). This is not surprising since CNL texts have to be clear and objective and should not include personal evaluations or opinions, which is precisely the kind of meaning typically codified in epistemic, evidential and illocutionary clausal adjuncts.

In the case of core/level one peripheries (total = 4,758 examples), we have found instantiations in all the positions in the LSC in the case of both contingency and temporal adjuncts: four positions in process adjuncts (LDP, medial, final, RDP), three positions in space adjuncts (LDP, medial, final) and two positions in pace adjuncts (medial, final).

The frequency results show (figure 3) that core/L1 peripheral constituents tend to appear naturally in final position in almost 60% of cases, but they are also frequently used in medial positions (24.79%). LDP is also seen in a smaller number of cases (15.26%). However, these constituents are less frequently located in clause initial positions (0.90%) and RDP (0.93%).

Figure 3. Positional preferences of core/level one adjuncts in the Airbus corpus



In table 3 we present the scale of positional preferences for core/L1 adjuncts in Simplified Technical English (ASD-STE100) in the Airbus corpus, together with examples and compare it with the same scale for Natural English, which shows a very similar array of positional preferences. The abbreviations used in the scales for Natural English represent the basic positions and their subvariants, as presented in Cortés-

Rodríguez and Rodríguez-Juárez (2019, 69): PERIni: Initial-position PERiphery; PERIniE: Initial-position PERiphery (end variant); PERMed: Medial-position PERiphery; PERMedE: Medial-position PERiphery (end variant); PERFinI: Final-position PERiphery (initial variant); PERFin: Final-position PERiphery; LDP/RDP: Left/Right Detached Positions.

Table 3. Core/L1 adjuncts: Scale of positional preferences

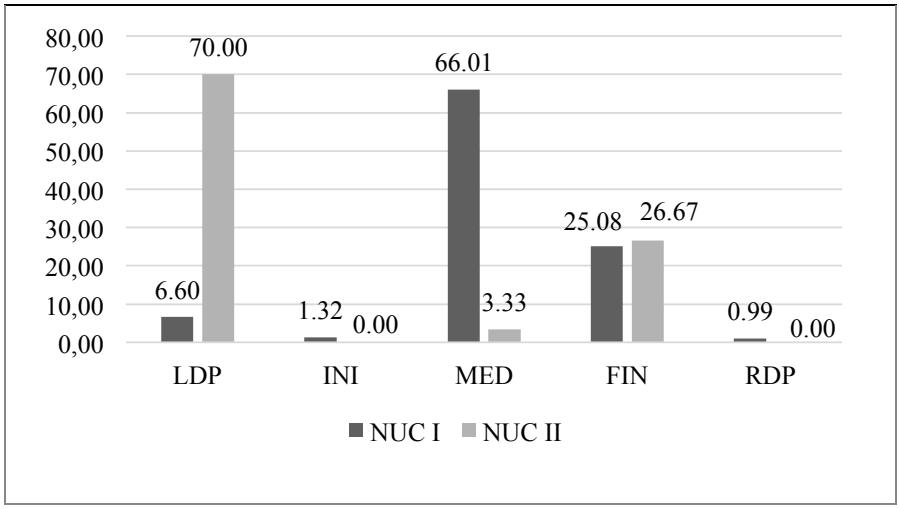
Airbus ASD-STE100	**FIN > MED > LDP > INI / RDP** Lower the trailing arm (25) *carefully with your hands* (FIN). *Slowly* (MED) close the FWD NLG door (31) with your hands. *Before the connection of the hose to the hydraulic filling valve* (LDP), make sure … *When the aircraft is on-ground* (INI) the MLG shock absorber is compressed. The ADCN interchanges data of the Normal Braking System (*through the BACS software*) (RDP).
Natural English	**PERFin > PERMed / LDP > PERIni** (Cortés-Rodríguez and Rodríguez-Juárez 2019, 73)

The group of nuclear peripheral elements is quite complex with respect to their positional behaviour and two tendencies that favour their representation into two groups of nuclear adjuncts can be observed (Cortés-Rodríguez and Rodríguez-Juárez 2019, 73):

- Nuclear periphery (I): Formed by focusing, degree and indefinite frequency group two—zero/low adverbs—and group three adjuncts—continual/universal. These adjuncts tend to be more rigid and occupy medial positions, although they can also be placed in final or initial clause internal positions.
- Nuclear periphery (II): comprised by definite frequency and indefinite frequency group one adjuncts (*usual*, *high* and *mid frequency*). These adjuncts tend to appear in final positions although the indefinite group one can frequently be found in other positions.

The results for the first group of nuclear adjuncts (NUC I; total = 301 examples) in the Airbus corpus corroborate their tendency to be placed in medial positions with 66.01% of the examples (see figure 4).

Figure 4. Positional preferences of nuclear adjuncts, types one and two in the Airbus corpus (total = 421 examples)



They are also common in final positions (25.08%), like the degree adjunct *too much* in the example given in table 4. Instances of *also* placed in LDP were also registered as well as a few examples of nuclear adjuncts, like *only*, placed in initial positions (1.32%).

Table 4. Nuclear Periphery (I): scale of positional preferences

Airbus ASD-STE100	**MED > FIN > LDP > INI > RDP**
	The bolts can become loose and *even* (MED) fall.
	Damage to the harness can occur if you bend it *too much* (FIN).
	Also (LDP), it has one restrictor and one flow control.
	The hitch pin *only* (INI) can be installed…
	This type of equipment (…) can cause damage to equipment, *specially to: electrical equipment …* (RDP)
Natural English	**PERMed > PERFin / PERIni** (Cortés-Rodríguez and Rodríguez-Juárez 2019, 73)

The comparison of the positional preferences of nuclear peripheries (I) both in the Airbus corpus and in Natural English reveals that in the CNL we registered examples of nuclear adjuncts in the two detached extra-clausal positions, although in low numbers, whereas in Natural English these positions were not recorded.

As for the second group of nuclear adjuncts (NUC II in figure 4), only three positions were registered (total = 120 examples). Their tendency in the Airbus corpus is to be located in LDP (70%), since there is a large number of examples of *usually* and *in general* placed in this position. However, they are also often located, like in Natural English, in final positions (26.67%; see table 5 for examples). Only four examples (3.33%) were registered in medial position: *usually* (3 occurrences) and *normally* (1 occurrence).

Table 5. Nuclear Periphery (II): scale of positional preferences

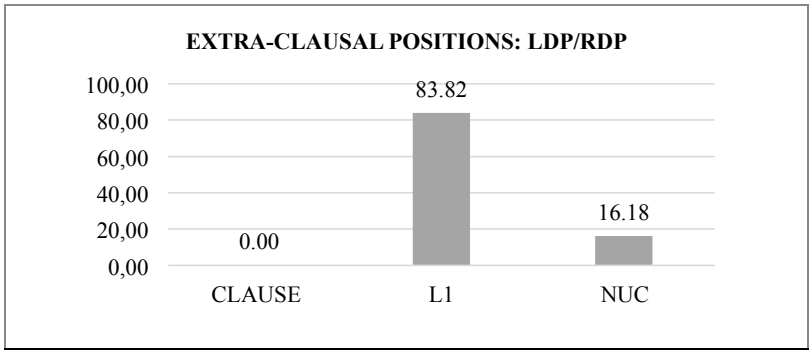
Airbus ASD-STE100	**LDP > FIN > MED** *Usually* (LDP), the lower set of seals keeps the pressure ... *In general* (LDP), these materials are flammable, poisonous and skin irritants. Do the test *five or six times* (FIN) ... This valve is *normally* (MED) de-energized and closed.
Natural English	**PERFin > PERIni / LDP / PERMed** (Cortés-Rodríguez and Rodríguez-Juárez 2019, 73)

The data from the two groups of nuclear peripheries justify the need to separate them into two distinct groups according to their positional behaviour, since the medial position is typically occupied by the first group of nuclear adjuncts (66.01%, figure 4) as opposed to the second, which is often placed in the LDP (70%) or final position (26.67%).

5.2. Peripheral Preference Depending on the Position in the LSC in the Airbus Corpus
In this section, we present the results from the point of view of the positions within the LSC, that is, what type of adjuncts are more often found in each of the positions in the LSC and we will compare the Airbus results with the results obtained in Natural English.

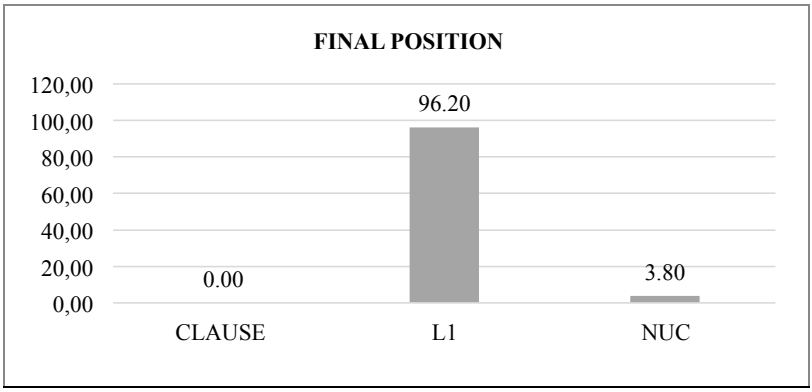
Let us begin with the extra-clausal positions. Core/L1 adverbials are much more frequent than nuclear adjuncts in both LDP and RDP in the Airbus corpus (see figure 5). In Natural English, however, LDP and RDP seem to be equally occupied by all types of peripheries for pragmatic purposes (Cortés-Rodríguez and Rodríguez-Juárez 2019, 73-74). So, the first notable difference between ASD-STE100 controlled language and Natural English is that, in the former, no examples of a clausal adjunct occupying either of these two detached positions were found. In Natural English, examples of both clausal illocutionary adjuncts ("Briefly (LDP), there is nothing more I can do about it," Quirk et al. 1985, 616) and clausal evidential adjuncts ("In my experience (LDP), such questions are seldom solved," Dik 1997a, 297) in the LDP were registered, as described in Cortés-Rodríguez and Rodríguez-Juárez (2019, 70).

FIGURE 5. Scale of adjunct preferences for the extra-clausal positions (LDP/RDP) in the Airbus corpus (RDP = 47 examples; LDP = 830 examples)



The final position is the predominant position for level one adjuncts in the Airbus corpus (96.2%; see figure 6), exactly as happens in Natural English, where they have been registered as showing a high preference (see table 6 and Cortés-Rodríguez and Rodríguez-Juárez 2019, 74), but it is rare to find nuclear adjuncts in this final position both in the CNL (3.8%) and in Natural English, where they show a medial preference (see table 6 and Cortés-Rodríguez and Rodríguez-Juárez 2019, 74). No examples of clausal adjuncts were registered in this position, neither in the Airbus corpus nor in Natural English.

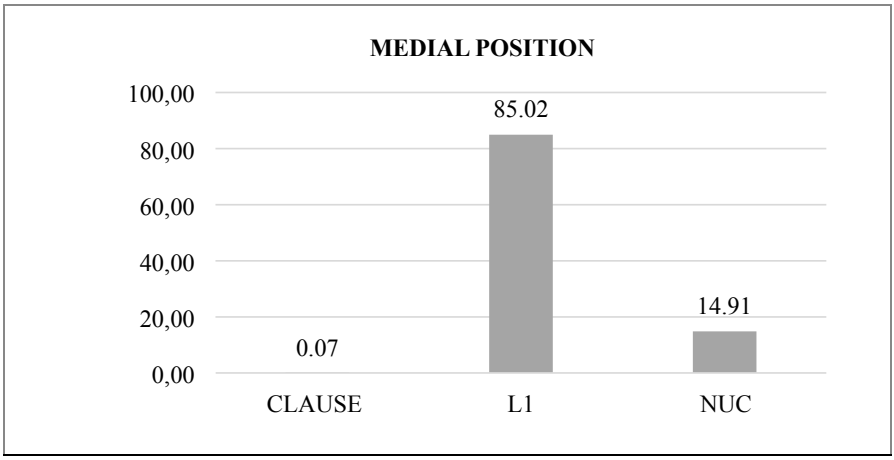
FIGURE 6. Scale of adjunct preferences for the final position in the Airbus corpus (2,872 examples)



As regards the medial position, 85.19% (figure 7) of the adjuncts located in this position in the Airbus corpus are examples of core/L1 adjuncts (*Slowly close the FWD NLG door with your hand; Make sure that you correctly align the Wheel with the axe; The*

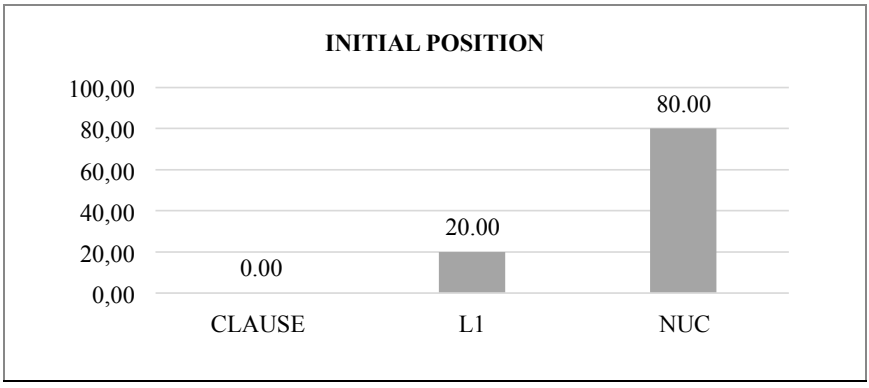
Normal Braking System is usually operated through the brake pedals). In Natural English, however, nuclear adjuncts show a higher preference for occupying this position than core/L1 adjuncts, which show a medial preference (see table 6 and Cortés-Rodríguez and Rodríguez-Juárez 2019, 74).

Figure 7. Scale of adjunct preferences for the medial position in the Airbus corpus (1,384 examples)



Finally, as we have already observed, the initial position is not very popular (only 47 examples), but it tends to be occupied by nuclear adjuncts (80%) in the Airbus corpus (figure 8). In Natural English, however, this position is more likely to be occupied by clausal adjuncts, followed by nuclear adjuncts (medial preference) and L1 peripheries (low preference; see table 6 and Cortés-Rodríguez and Rodríguez-Juárez 2019, 74).

Figure 8. Scale of adjunct preferences for the initial position in the Airbus corpus (47 examples)



In general, then, we can conclude that, in the Airbus corpus, the position that is more often occupied by peripheral constituents is the final position (55.44% of the total of adjuncts analysed in the sample corpus), followed by the medial position and the LDP and, on a much smaller scale of frequency, by the initial position and the RDP (figure 9).

Figure 9. Positional preferences in the Airbus corpus

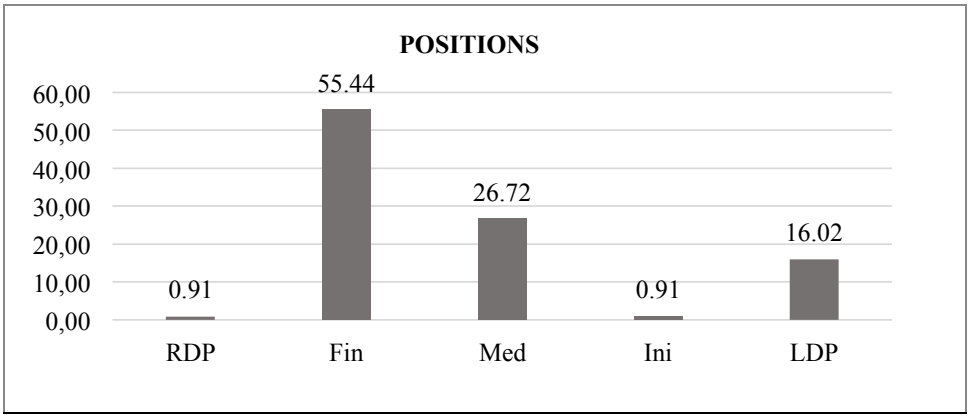


Table 6 below shows all the scales of positional preferences that have been postulated by Cortés-Rodríguez and Rodríguez-Juárez (2019, 74) for Natural English, where examples of all the peripheries in all the positions with different levels of positional preferences were registered, except for clausal adjuncts in final position.

Table 6. Scale of positions and peripheral preferences in Natural English (Cortés-Rodríguez and Rodríguez-Juárez 2019, 74)

		POSITIONS						
		LDP	*PER Ini*	*PERIniE*	*PERMed*	*PERMedE*	*PERFinI*	*PERFin*	RDP
PERIPHERAL PREFERENCES	***+High***	all	PER_CL		PER_NUC		PER_CONSTRL1		all
	± Mid		PER_NUC		PER_CONSTRL1		PER_NUC		
	−Low		PER_CONSTRL1		(PER_CL)				

Table 7 shows the scale of positions and peripheral preferences that Cortés-Rodríguez and Rodríguez-Juárez (2019, 76) provided in their preliminary study of a small sample of the Airbus corpus, along with the non-occurrences that were observed with respect to Natural English. Here, examples of clausal adjuncts in the LDP and RDP were not

found, nor in the initial and medial positions; as regards core/L1 adjuncts, no examples were found in either initial or medial positions.

Table 7. Changes on the scale of positions and peripheral preferences in the study of a reduced subcorpus of Airbus ASD-STE100 (Cortés-Rodríguez and Rodríguez-Juárez 2019, 76)

<table>
<tr><th colspan="2" rowspan="2"></th><th colspan="7">POSITIONS</th></tr>
<tr><th>LDP</th><th>PER Ini</th><th>PER IniE</th><th>PER Med</th><th>PER MedE</th><th>PER FinI</th><th>PER Fin</th><th>RDP</th></tr>
<tr><td rowspan="3">PERIPHERAL PREFERENCES</td><td>+High</td><td rowspan="3">PER_NUC
PER_ CONSTRL1
<s>PER_CL</s></td><td colspan="2"><s>PER_CL</s></td><td colspan="2">PER_NUC</td><td colspan="2">PER_ CONSTRL1</td><td rowspan="3">PER_NUC
PER_ CONSTRL 1
<s>PER_CL</s></td></tr>
<tr><td>± Mid</td><td colspan="2">PER_NUC</td><td colspan="2"><s>PER_ CONSTRL1</s></td><td colspan="2">PER_NUC</td></tr>
<tr><td>−Low</td><td colspan="2"><s>PER_ CONSTRL1</s></td><td colspan="2"><s>(PER_CL)</s></td><td colspan="2"></td></tr>
</table>

In the present study, where we have expanded our analysis to the whole corpus of Airbus adverbials, we have been able to demonstrate that the number of peripheries depending on the position in the LSC is more reduced than Cortés-Rodríguez and Rodríguez-Juárez (2019, 76) predicted, since we have found instances of core/L1 adjuncts in initial position with a low frequency (43 examples, 0.90%; see table 8) and a moderate number of examples (1,179) in medial position (24.79%). We have also observed that the preference for nuclear adjuncts to occur in initial position is low (0.95%; table 8) and not medium, as Cortés-Rodríguez and Rodríguez-Juárez previously suggested (2019, 74). Table 8 illustrates the differences in positional preferences depending on the type of peripheral constituent:

Table 8. Positions and peripheral preferences in the complete Airbus corpus

PERIPHERAL PREFERENCES	POSITIONS				
	LDP	INITIAL	MEDIAL	FINAL	RDP
+High (100%-51%)				CORE/L1 (58.12%)	
± Medial (50%-20%)	NUC (29.59%)		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%)

The findings on the frequency distribution with respect to the position of Airbus adjuncts has allowed us to assign a weight with regard to the predictability of the types of adjuncts that can occupy the different positions. Each position will be given a specific value based on the analysis of the Airbus positional preferences and frequencies that we have established. 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. Table 9 shows the weights that have been given to each position and for each type of adjuncts following the different rates of frequency registered in the Airbus corpus.

Table 9. Assignment of weight to the different positions depending on the adjunct type based on frequency rates in the Airbus corpus

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

WEIGHT	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%

In other words, the preferences of Airbus peripheral constituents depending on the position in the LSC are accounted for by means of the weight-based priority established in accordance with the probabilistic frequency registered in our corpus-based research for each type of adjunct in each position and which could be used in the computational parsing of Airbus adverbials in ARTEMIS.

6. Conclusions

In this study we have offered an exhaustive analysis of adverbials in the CNL ASD-STE100 of the texts that form the Airbus corpus. Following the classification of these types of constituents in RRG, we have analysed the positional preferences of the two types of adjuncts that are found in the corpus (nuclear and core/L1 peripheries). In our analysis we have also drawn a comparison between the behaviour of adverbials in this CNL and Natural English. This has revealed some differences that are probably due to the restricted nature of ASD-STE100, in which, for example, there are no clausal

adjuncts in any of the positions—except for a single example in medial position. Furthermore, the data obtained have been contrasted with a previous more restrictive study and our findings in the current work provide a more refined overview of the positional preferences of adverbials in the corpus, which has led us to establish a scale of markedness (weight) of the different types of adjuncts depending on the position in the clause.

The establishment of a weight factor is especially relevant for the automatic processing of adverbials in general, and more particularly for the computational parsing of these constituents in any documentation produced in the controlled language ASD-STE100, such as the texts of the Airbus corpus. The semantic variability and positional flexibility of adjuncts present a challenge for any syntactic processing by computational means. We believe that this weight factor can significantly increase the accuracy of any NLP resource designed for the retrieval of an appropriate syntactic and semantic representation of linguistic expressions. In this regard, some advances have already been made in the design of the parsing rules and the AVMs that are needed for the automatic parsing of Airbus adverbials in the prototype ARTEMIS (Cortés-Rodríguez and Rodríguez-Juárez, in press). However, as was mentioned in section one, since ARTEMIS is still a prototype in the testing phase, the results obtained for adverbials in Natural English in our research will be of relevance in a later stage of implementation of this parser.

Works Cited

Aerospace and Defence Industries Association of Europe. 2017. *ASD-STE Simplified Technical English Specification STE-100. International Specification for the Preparation of Technical Documentation in a Controlled Language*. Issue 7. Brussels.

Anthony, Laurence, comp. 2019. *AntConc* (version 3.5.9). Tokyo: Waseda University.

Blecua, José Manuel et al., eds. 1999. *Filología e informática. Nuevas tecnologías en los estudios filológicos*. Barcelona: Milenio and Universidad Autónoma de Barcelona.

Cortés Rodríguez, Francisco J. and Carolina Rodríguez-Juárez. 2019. "The syntactic parsing of ASD-STE100 adverbials in ARTEMIS." *Revista de Lingüística y Lenguas Aplicadas* 14 (1): 59-79.

—. (In press). "Computational Analysis of Adjuncts in ASD-STE100 for the NLP Parser ARTEMIS." *Vigo International Journal of Applied Linguistics (VIAL)*.

Díaz-Jorge, Victoria. 2017. "Adjuncts in Role and Reference Grammar: The Peripheries." BA thesis, Univerity of La Laguna.

Dik, Simon C. 1997a. *The Theory of Functional Grammar. Part 1: The Structure of the Clause*. In Hengeveld (series editor) 1997.

—. 1997b. *The Theory of Functional Grammar. Part 2: Complex and Derived Constructions*. In Hengeveld (series editor) 1997.

Downing, Angela and Philip Locke. 1992. *A University Course in English Grammar*. London: Prentice Hall.

Felices-Lago, Ángel and Ángela Alameda-Hernández. 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.

Fumero Pérez, Mª Carmen and Ana Díaz-Galán. 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: 149-74.

García Ferrando, Manuel. 1985. *Introducción a la estadística en sociología*. Madrid: Alianza Universidad.

Hengeveld, Kees, ed. 1997. *The Theory of Functional Grammar*. 2 vols. Berlín: De Gruyter.

— and J. Lachlan Mackenzie. 2008. *Functional Discourse Grammar: A Typologically-Based Theory of Language Structure*. Oxford: Oxford UP.

Hengeveld, Kees et al. 2019. "Perception Verbs in Brazilian Portuguese: A Functional Approach." *Open Linguistics* 5: 268-310.

Huddleston, Rodney and Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge UP.

Kuhn, Tobias. 2014. "A Survey and Classification of Controlled Natural Languages." *Computational Linguistics* 40 (1): 121-70.

Martín-Díaz, Mª Auxiliadora and Marta González-Orta. 2020. "Clausal Arguments and Peripheries in ASD-STE100: The Parsing of Subordination in ARTEMIS." *Revista Electrónica de Lingüística Aplicada* 19 (1): 1-23.

Neuber, Keith A. 1980. *Needs Assessment. A Model for Community Planning*. Newbury Park: SAGE.

Nolan, Brian and Elke Diedrichsen, eds. 2013. *Linking Constructions into Functional Linguistics: The Role of Constructions in Grammar*. Amsterdam and Philadelphia: John Benjamins.

— and Carlos Periñán-Pascual, eds. 2014. *Language Processing and Grammars: The Role of Functionally Oriented Computational Models*. Amsterdam and Philadelphia: John Benjamins.

Periñán-Pascual, Carlos. 2012. "En defensa del procesamiento del lenguaje natural fundamentado en la lingüística teórica." *Onomázein* 26: 13-48.

—. 2013. "A Knowledge-Engineering Approach to the Cognitive Categorization of Lexical Meaning." *VIAL: Vigo International Journal of Applied Linguistics* 10: 85-104.

— and Francisco Arcas-Túnez. 2014. "The Implementation of the FunGramKB CLS Constructor." In Nolan and Periñán-Pascual 2014, 165-96.

Quirk, Randolph et al. 1985. *A Comprehensive Grammar of the English Language*. New York: Longman.

Ruiz de Mendoza, Francisco. 2013. "Meaning Construction, Meaning Interpretation and Formal Expression in the Lexical Constructional Model." In Nolan and Diedrichsen 2013, 231-70.

— and Alicia Galera Masegosa. 2014. *Cognitive Modeling. A Linguistic Perspective*. Amsterdam and Philadelphia: John Benjamins.

Spaggiari, Laurent, Florence Beaujard and Emmanuelle Canneson. 2005. "A Controlled Language at Airbus." *Lingusitcae Investigationes* 28 (1): 107-22.

Van Valin, Robert D. 2005. *Exploring the Syntax-Semantics Interface*. Cambridge: Cambridge UP.

— and Randy J. LaPolla. 1997. *Syntax*. *Structure*, *Meaning and Function*. Cambridge: Cambridge UP.

Received 19 July 2021 Revised version accepted 22 November 2021

Carolina Rodríguez-Juárez works at the Department of Modern Philology, Translation & Interpretation, Universidad de Las Palmas de Gran Canaria, Spain. Her research focuses on the interaction between lexis & grammar within the theory of Role and Reference Grammar; the computational treatment of the lexicon and constructional meaning in FunGramKB; and the development and implementation of tools designed for natural language processing.

Francisco José Cortés-Rodríguez is Full Professor of English Language and Linguistics at the University of La Laguna, Spain. He has been the Head of the 'Andrés Bello' Institute for Linguistics and is currently the President of AESLA (Spanish Society of Applied Linguistics). His areas of research interests are word-formation, lexicology, the grammar-lexis interface within functional and cognitive models and computational linguistics.