

2022-09-15

When Articles Go Missing: Analyzing Optionality in Spanish Prepositional Relative Clauses

Levinstein Rodriguez, Andrea

Levinstein Rodriguez, A. (2022). When articles go missing: analyzing optionality in Spanish prepositional relative clauses (Master's thesis, University of Calgary, Calgary, Canada).

Retrieved from <https://prism.ucalgary.ca>.

<http://hdl.handle.net/1880/115259>

Downloaded from PRISM Repository, University of Calgary

UNIVERSITY OF CALGARY

When Articles Go Missing: Analyzing Optionality in Spanish Prepositional Relative Clauses

by

Andrea Levinstein Rodriguez

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF ARTS

GRADUATE PROGRAM IN LINGUISTICS

CALGARY, ALBERTA

SEPTEMBER, 2022

© Andrea Levinstein Rodriguez 2022

ABSTRACT

This thesis project presents a comprehensive overview of article optionality in Spanish prepositional relative clauses. This definite article, which is part of the operator phrase, is subject to a complex pattern of variation that regulates whether it can be present or absent in a given phrase. Through a mixture of experimental and formal analysis, this project investigates (i) the syntactic and semantic factors that influence the presence or absence of the article and (ii) the role of the article within the syntactic structure of Spanish prepositional relative clauses.

A corpus study found that the article was significantly more likely to be absent when the antecedent of the relative clause was definite, inanimate, or plural, or when the embedded clause was not negated. In addition to these factors, a significant effect of preposition and an interaction between definiteness and the preposition *en*, which had not been attested in the literature, were found. Subsequently, an acceptability judgement task conducted on eighty-eight Mexican Spanish speakers showed higher acceptability ratings for article presence, definite antecedent, and the preposition *en*. The *en*:Definiteness interaction was replicated in post-hoc testing.

Based on these results, I present a modified version of Cinque's (2020) double-headed structure. I propose that prepositional relative clauses without the article follow Cinque's matching analysis-like derivation, while those with the article contain a larger relative operator with a [+specific] feature, thus avoiding deletion. This project presents an updated understanding of an under-researched area of Spanish grammar. It provides empirical support to the claim that the presence or absence of the article is not a matter of free variation, and makes headway on identifying the factors that mediate the presence or absence of the article, as well as the mechanisms that underlie them.

Table of Contents

ABSTRACT	ii
PREFACE	v
LIST OF SYMBOLS, ABBREVIATIONS, & NOMENCLATURE	vi
ACKNOWLEDGEMENTS.....	vii
1. INTRODUCTION.....	1
2. LITERATURE REVIEW	4
2.1. General Literature on Prepositional Relative Clauses.....	4
2.1.1. What is a relative clause?.....	4
2.1.2. Semantic types.....	5
2.1.3. Prepositional RCs	6
2.2. Literature on Prepositional RCs in Spanish	8
2.2.1. Semilibre Relative Clauses.....	8
2.2.2. Descriptive accounts of article optionality.....	9
2.2.3. Unique que hypothesis.....	11
2.2.4. Bare Que hypothesis	13
2.2.5. Conclusions on the categorization of ‘que’	16
2.3. Current analyses of prep-RCs.....	17
2.3.1. Pied-piping.....	18
2.3.2. Head-Antecedent relationship	22
2.4. Empirical study of Article Optionality in prep-RCs	25
2.4.1. Summary of Santana Marrero (2004).....	25
2.4.2. Critique of Santana Marrero (2004)	27
3. CORPUS ANALYSIS.....	29
3.1. Methodology	29
3.2. Coding and Classifying	33
3.3. Statistical report.....	38
3.4. Discussion	42
3.4.1. Summary of Results.....	42
3.4.2. What can we do with the preposition effect?.....	43
4. ACCEPTABILITY JUDGEMENT TASK	45
4.1. Study Design	45
4.1.1. Variables.....	45

4.1.2.	Predictions	48
4.1.3.	Lists.....	48
4.1.4.	Procedure.....	49
4.2.	Participants.....	50
4.3.	Results	51
4.3.1.	Lmer	52
4.3.2.	Initial post-hoc testing	53
4.3.3.	Analysis of between-participant differences.....	56
4.3.4.	Lmer modeling with random slopes	57
4.4.	Preliminary Discussion	58
5.	DISCUSSION.....	60
5.1.	Overview of results	60
5.2.	Grounding the definiteness data in the theory	61
5.2.1.	Brucart and the definiteness effect.....	62
5.2.2.	Updating my analysis through Cinque (2020).....	62
5.3.	A structural account of article alternation	64
5.4.	Outstanding issues	66
5.5.	Can we ground the preposition effect in the theory?	68
6.	CONCLUSION	71
	REFERENCES.....	75
	APPENDIX.....	80

PREFACE

This thesis is original, unpublished, independent work by the author, Andrea Levinstein Rodriguez. The study reported in Chapter 4 was covered by Ethics Certificate REB21-0606, issued by the University of Calgary Conjoint Health Ethics Board for the project “Optionality in Spanish Prepositional Relative Clauses” on September 24, 2021.

LIST OF SYMBOLS, ABBREVIATIONS, & NOMENCLATURE

ART	definite article
PREP	preposition
OP	(null) relative operator
1	first person
2	second person
3	third person
SG	singular grammatical number
PL	plural grammatical number
F	feminine grammatical gender
M	masculine grammatical gender
N	neuter grammatical gender
ACC	accusative marker
DAT	dative marker
OBJ	object clitic
CL	clitic (unspecified)
DIST	distal pronoun, 'that, those'
PROX	proximal pronoun 'this, these'
SJV	subjunctive mood
COND	conditional mood
EXT	existential (in the context of <i>be</i> , 'there is/are')
PRT	participle
PSV	passive marker
FUT	future tense
PST	past tense
INF	infinitive
ERG	ergative
REL	relativizer
REFL	reflexive

ACKNOWLEDGEMENTS

There are many people without whom this project would have never happened and who have been vital to my academic journey. First and foremost, my supervisor, Dr. Dennis Storoshenko, has been a role model and an inspiration from the moment I met him. The ease with which you share your knowledge and your unwavering support have been invaluable. I cannot imagine having a better supervisor. I must also express my deepest gratitude to the other members of my examination committee: Dr. Betsy Ritter, Dr. Maria Cristina Cuervo, and neutral chair Dr. Dimitrios Skordos, who generously gave me their time, knowledge, and expertise. You have really expanded my understanding of this project's potential. Additionally, I must acknowledge Dr. Ritter in her role as my graduate advisor. She saw my potential long before I did and helped me realize that grad school could be more than a pipe dream. Truly, I could not have done it without you.

I would also like to thank Dr. Susanne Carroll, who gave me my first taste of lab research and supervised my undergraduate thesis. Not only did that project inspire the present work, but her clear mind and scientific rigor inspire me as an academic. Likewise, Dr. Lindsay Hracz has been a constant source of support, mentorship, and inspiration since my undergrad days.

The University of Calgary's linguistics department is an incredibly supportive environment that has really allowed me to flourish as an academic. The faculty members have all contributed in some way or another to my development as a graduate student and provided thoughtful feedback on my project. Particular thanks go to Dr. Steve Winters, without whom I would still be scared of statistical modeling. I have also received such support and warmth from my peers, especially the members of the Experimental Syntax Lab, who have seen this thesis come together from the beginning and provided thoughtful feedback and support at every step. I am also grateful to the members of A Higher Clause, the executive members of Verbatim, and any classmates not covered by those two groups. And of course, I must not forget everybody who participated in the online task and/or shared the recruitment poster. Their warmth and enthusiasm made this such a rewarding experience and affirmed my love for experimental linguistics.

Finally, I have to thank all the people in my life who have encouraged me, supported me, and kept me relatively sane. My community at Temple B'nai Tikvah has shown me such

support and enthusiasm for my success, and been there for me in tough times. I must single out Caron Glickman, who has been the truest of friends. My D&D group has been a refuge, preventing me from turning this thesis into my entire personality. Thank you also to all my friends, who have put up with my disappearing acts. Finally, I must thank my family. I inherited my mom's relentless spirit and unwillingness to give up; it is thanks to her steadfast support that I can be here today. I cannot even begin to list all the ways in which Nadine, my younger sibling, has been there for me. Their willingness to always lend a hand or an ear and their unwavering faith in me have propelled me throughout this journey.

1. INTRODUCTION

This research project is centered on the study of the prepositional relative clause (henceforth prep-RC) in Spanish, specifically those clauses that are headed by the complementizer *que*. These prep-RCs are characterized by the presence of a nominal particle that is homophonous with the definite article *el/la/los/las*, expressing the same number and gender features as the antecedent of the relative clause.

- (1)¹ La mesa [CP en (**la**) que puse el libro]
 the.F,SG table on ART.F,SG QUE put.1SG the book
'The table on which I put the book'

This particle (henceforth 'the article') exhibits a complex pattern of behaviour in terms of the circumstances under which it can be present or absent within the prep-RC. This complexity is often discussed in terms of optionality, but some factors seem to fully prevent the article from being absent. For example, the article can only be absent when preceded by a select number of mostly monosyllabic prepositions, although the exact list of valid prepositions varies throughout the literature. We can see this in the contrast between examples (1) and (2), which are preceded by monosyllabic *en* 'on' and bi-syllabic *ante* 'in front', respectively. Other factors, e.g., certain properties of the antecedent, are better described as having a gradient effect. That is, these factors predispose the likelihood of the article being present or absent, but exceptions and interpersonal variation are common.²

- (2) *El edificio ante que nos encontramos es de estilo Barroco.
 the.M,SG building before QUE REFL,1SG found.1PL is of style baroque
'The building in front of which we found ourselves is of (the) baroque style.'

(Porto Dapena, 1997, p. 24)

My research project aims to provide a comprehensive overview of the factors that influence the presence or absence of this article, and to deepen our syntactic understanding of Spanish

¹ Unless otherwise specified, all example sentences and translations from Spanish sources are my own. Phi features (number and gender) are only specified when relevant.

² It is worth acknowledging that, while this thesis implicitly treats the presence of the article as the default state of prep-RCs, this is done for simplicity's sake and not as a theoretical claim. Whether either state of the article could be considered the more standard form is beyond the scope of this paper.

prep-RCs, specifically those aspects pertaining the article. For the first goal, I hope to provide empirical backing to the generalizations that have been proposed in the literature, to advance our collective understanding of which factors are mandatory and which are not, and to develop methods for how to distinguish between the effects of these factors. For the second goal, I want to untangle the assumptions that are made in the literature about the article: is it best characterized as a determiner, as is standard for articles, or could it be, for example, a pronominal clitic or an agreement marker? Should we be talking about its alternation in terms of presence vs. absence, or could it be a distinction of null vs. overt, or a reflection of divergent syntactic structures? What is its relationship to the rest of the embedded clause?

Moreover, my project hopes to contribute to the literature by bringing attention to a linguistic object (the article) that is often overlooked. Research into article optionality in prep-RCs is quite limited, and often presented as part of a broader research program – from syntactic proposals about Spanish RCs (Plann, 1980; Brucart 1992, 1994, 1999; Suñer, 2001) to analyses of diachronic variation (Guzmán Riverón, 2012; Blas Arroyo & Vellón Lahoz, 2017, 2018; Vellón Lahoz, 2019, 2020). This project is one of a very small number to make the article its primary object of research, and the first to use experimental syntax methods to accomplish this goal.

There are two facets to the data-gathering aspect of this project: a corpus study and an acceptability judgement task. These studies provide positive and negative data, respectively, on the contexts in which it is (un)acceptable for the article to be absent in a prep-RC. The corpus study is a broad exploration that analyses the factors that have been identified in the literature as influencing the presence/absence of the article. The corpus identified which factors and interactions between factors are the most promising avenues of deeper investigation. The corpus found that the presence or absence of the article significantly correlated with the Definiteness, Animacy and Number of the antecedent, as well as with Preposition and Sentence Polarity. I chose Preposition and Definiteness as the most promising factors to carry forward. I hypothesized that Preposition reflected an underlying distinction between adjunct PPs and argument PPs. With these factors in mind, I designed and implemented an Acceptability Judgement Task (AJT). Although the argument/adjunct

distinction was not significant, the AJT found further evidence of the significant effects of Preposition and Definiteness on the acceptability of sentences without the article. The data gathered in these tasks was then integrated into our theoretical understanding of prep-RCs, serving as the basis for an updated understanding of the structure of Spanish prep-RCs, in which the presence or absence of the article is motivated by a difference in the structure of the relative operator.

The structure of this thesis is as follows: Chapter two (re-)introduces our object of study and provides an overview of the Spanish syntactic literature and current research into this topic. In this chapter, I make some initial choices about the theoretical assumptions and frameworks that will be incorporated into my analyses. Chapter three contains the corpus study – the design, implementation, and statistical analysis. I explain how the results to develop the hypotheses that I would carry onto my main experimental task. In Chapter four, I go over the design, recruitment, and results of the acceptability judgement task. Chapter five examines the results of the previous two tasks in the context of what they can contribute to a syntactic analysis of prepositional RCs. In this chapter, I present a formal syntactic analysis of what, based on my observations, I propose the structures of prep-RCs with and without the article to be. Chapter six is the conclusion.

2. LITERATURE REVIEW

This research project aims to answer **(i)** what are the factors that influence the presence/absence of the optional article in prepositional relative clauses (prep-RCs) in Spanish, and **(ii)** what role does this article play in the syntactic structure of these phrases. This section gives an overview of the relevant literature on prepositional relative clauses. Section 2.1 (re-)introduces the object of study and briefly goes over some of the primary distinctions one must keep in mind when studying prep-RCs. Section 2.2 delves into the study of prep-RCs in Spanish, focusing on those accounts that explicitly discuss the article. Section 2.3 introduces contemporary generativist frameworks that might account for some aspects of prep-RCs that are under-discussed in the Spanish literature. Section 2.4 presents the existing empirical research into article optionality and motivates the need to gather and analyse new data.

2.1. General Literature on Prepositional Relative Clauses

2.1.1. What is a relative clause?

Relative clauses are a type of embedded clause that modify a nominal. The modified nominal, known as an antecedent (*el libro* in (3)), is interpreted as being part of both the main and the embedded clause. In (3), the embedded clause is interpreted as *puse el libro en la mesa* ‘I put the book on the table’. This effect is standardly assumed to be the result of co-indexation with a Relative Operator/*wh*-word (null in (3)). The Operator phrase merges as part of the embedded clause (in (3), it is the Direct Object of *puse* ‘I put’) and moves to Spec, CP through *wh*-movement. This movement is motivated by a syntactic feature in the Complementizer (*que* ‘that’ in (3)).

- (3) Voy a leer el libro_i [CP \emptyset_{1i} que puse \emptyset_{1i} en la mesa]
 go.1SG to read the book OP that put.1SG on the table
‘I am going to read the book that I put on the table’

Although both the Complementizer and the Operator are crucial components of *wh*-movement, it is often the case that one of the components is not phonologically overt (Horvath, 2017). Within generative frameworks, it is standardly assumed that both components are present even when one of them is null. In descriptive studies, both the

Operator and the complementizer are often lumped under the descriptive category of “relative heads” or “relative pronouns”.^{3,4} In Spanish, this ambiguity is at the heart of one of the major topics in the study of prep-RCs: whether *que* in this structure should be categorized as a complementizer (analogous to ‘that’) or an operator (analogous to ‘which’).⁵ This debate and its consequences for my object of study are examined in detail in Section 2.2.

2.1.2. *Semantic types*

Another major factor that impacts our understanding of relative clauses is semantic classification. Linguists have traditionally distinguished between two mayor types of relative clauses, restrictive and appositive. Restrictive relative clauses specify a potential referent from within a group. In (4a), the relative clause singles out a specific table (the one where the book is) from a set of potential tables. Syntactically, they are traditionally represented as modifying a Noun Phrase (depicted in (4b)), although recent analyses (Wiltschko, 2012; Cinque, 2020, among others) often depict them as modifying an intermediate nominal phrase like *nP* or *NumP*. Appositive relative clauses provide additional information about a given referent. In (5a), the relative clause provides additional information about the wet table (that there is a book on top of it). It creates the pragmatic implication that the table being wet is a problem because the book might get damaged. Syntactically, they are standardly represented as modifying a Determiner Phrase (depicted in (5b)).

- (4) a. La mesa [CP en (la) que puse el libro] está mojada
 the table on ART.F.SG QUE put.1SG the book is wet
‘The table on which I put the book is wet’

- b. [DP la [NP mesa [RC en (la) que puse...]]]

³ These relative heads are subcategorized as pronominal ((ART+) *que*, *quien*, and ART+*cual*, equivalent to ‘which’ and ‘who’); possessive (*cuyo*, equivalent to ‘whose’); quantifier (*cuanto*); and adverbial (*donde*, *cuando*, *como*, equivalent to ‘where’, ‘when’ and ‘how’, respectively) (Lopez, 2010).

⁴ However, some authors (e.g., Suñer, 2001) use the term ‘relative pronoun’ to refer exclusively to a *wh*-word that is an operator.

⁵ Where not directly relevant, I will gloss *que* simply as QUE in all example sentences and avoid representing the null element.

- (5) a. La mesa, [CP en (la) que puse el libro], está mojada
 the table on ART.F.SG QUE put.1SG the book is wet
 ‘The table, on which I put the book, is wet’

b. [DP [DP la mesa]][RC en (la) que puse...]]

One of the most reliable diagnostics for differentiating between these two types of RCs is prosodic structure. Appositive RCs form a distinct prosodic phrase, while restrictive RCs are part of the prosodic structure of its antecedent.⁶ In writing, this is often represented through the use of commas or m-dashes, although orthography is not recommended as a diagnostic (Loock, 2010, pp. 12-14). The syntactic differences between these types of RCs have important consequences with regards to their distribution (what antecedents they can take, what tenses they can use, etc.). Since appositive RCs are characterized as having less constraints than restrictive RCs, most research on RCs tends to focus on restrictive RCs (Loock, 2010).

2.1.3. Prepositional RCs

The main focus of this research project is a sub-type of RCs known as prepositional relative clauses, in which the operator-phrase (*wh*-word *cual* in (6)) merges with a preposition (*en* in (6)) inside of the embedded clause. The resulting prepositional phrase moves to Spec, CP “carried” by the operator phrase. This phenomenon is known as pied-piping. While in English, it is possible for the preposition to “remain behind” (known as Preposition Stranding (7)), pied-piping is the primary method of prep-RC formation in Spanish.⁷ The mechanisms underlying pied-piping are discussed in section 2.3.1.

⁶ This distinction has been attested cross-linguistically (Loock, 2010), but, to the best of my knowledge, it has not specifically been investigated for Spanish, although Lopez (2010) attests to its existence in this language.

⁷ This is not to say that it is the sole mechanism available in Spanish, especially in cases involving *que*. Another relatively common mechanism is Resumption, in which a pronoun shows up in what is assumed to be the original location of the *wh*-word (i). A third, less common, mechanism is null-prep (also known as preposition chopping or *cortada*), in which a mandatory preposition is elided from the sentence (ii), giving it the appearance of a non-pied-piped RC (Lopez, 2010).

(i) Es una persona que yo sí confío en ella
 is.3SG a.F.SG person QUE I yes trust.1SG in her
 ‘She is a person that I do trust her.’

(Lopez 2010, p.2)

- (6) La mesa [CP [PP en la cual]₁ Ø_c puse el libro ~~en la cual~~]
 the table on ART.F.SG which put.1SG the book
'The table on which I put the book'

- (7) The table [CP which₁ I put the book on ~~which₁~~]

Specifically, I am interested in the particle that appears between the preposition and *que* in sentences like (8). This article has matching φ -features with the antecedent, e.g., *la* has the same 3rd person singular feminine features as '*la mesa*' in (8) below. As we will explore in the following section, evidence indicates that the presence or absence of the article is mediated by a number of factors, a phenomenon henceforth referred to as article optionality.⁸ In the literature, this particle is standardly labeled as an article, since it shares the agreement paradigm with the standard definite article (*el/la/los/las*, 'the').⁹ However, this claim is not uncontested. In the diachronic literature, where most of the research on article optionality is centered, the degree of grammaticalization that the article has undergone is subject to some debate, with some sources classifying it instead as a referential pronoun or an agreement affix (see Blas & Vellón, 2018, and sources within). In the next section, I discuss how article optionality has been discussed in the synchronic literature, focusing on the factors that have been proposed to influence optionality.

- (8) La mesa [CP en (la) que puse el libro]
 the table on ART.F.SG QUE put.1SG the book
'The table on which I put the book'

-
- (ii) Yo lo pongo en colegios Ø que hay pocos niños.
 I ACC.3SG.M put.1SG in schools QUE be.EXT few children
'I put him in schools (in) which there are few children.'

(Lopez 2010, p. 95)

⁸ Throughout this thesis, I will be using <PREP+ART+*que*> to refer specifically to prep-RCs where the article is present, and <PREP+*que*> to refer to prep-RCs where it is absent.

⁹ In addition to the article paradigm, the particle also appears as the so-called neuter article *lo*, which is used to refer to non-nouns (adjectives, propositions, etc.). Although this article is standardly characterized as having a neuter grammatical gender, recent literature (e.g., Stark & Pomino, 2010; Zulaica Hernandez, 2018) has argued that it is actually devoid of φ -features. This neuter *lo* (not to be mistaken with the M.SG. object clitic *lo*) is systematically excluded from discussions on the Spanish prep-RC and article alternation (but see: Plann, 1980). The specific properties and distribution of *lo* are beyond the scope of this thesis.

2.2. Literature on Prepositional RCs in Spanish

This section provides an overview of the syntactic literature that has focused on prepositional Relative Clauses in Spanish. Beyond some necessary background, this section focuses on those approaches that discuss article optionality.

Section 2.2.1 presents a background issue that consistently shows up in discussions of Spanish prep-RCs, that of homophony with the *semilibre* relative clause. Section 2.2.2 introduces some of the descriptive accounts of Spanish prep-RCs, with focus on their discussion of article optionality. Section 2.2.3 introduces Brucart's (1992) Unique Que hypothesis and how this theory contributes to the study of the article. Section 2.2.4 covers the Bare Que hypothesis, focusing on Suñer's (2001) objections to Brucart's work and her position on article optionality. Section 2.2.5 compares the Unique Que and Bare Que hypotheses and concludes that following the Unique Que hypothesis presents the more viable avenue for my research goals.

2.2.1. *Semilibre* Relative Clauses

One of the major issues surrounding the literature on prep-RCs is the importance of differentiating between a prepositional RC with the article present (<PREP+ART+*que*>) (9a) and the <ART+*que*> "relative pronoun," most prominently in cases where <ART+*que*> is preceded by a preposition (9b).

- (9) a. Recientemente, me gané una aspiradora [CP con la que
 recently CL.1SG won.1SG a.F.SG vacuum with ART.F.SG QUE
 voy a limpiar la alfombra]
 going.1SG to clean.INF the carpet
 'Recently, I won a vacuum with which I will clean the carpet.'
- b. En vez de la aspiradora vieja, limpié la alfombra
 instead of the.F.SG vacuum old.F.SG cleaned.1SG the carpet
 con la [CP que me gané recientemente]
 with ART.F.SG QUE CL.1SG won.1SG recently
 'Instead of the old vacuum, I cleaned the carpet with the (one) I won recently.'

As far back as Bello (1847, as cited in Suñer, 2001), researchers focused on Spanish syntax have noticed that these seemingly identical “relative pronouns” have different underlying structures. The most thorough of these approaches is Plann (1980), a work dedicated to differentiating between these two structures from a Transformational Grammar framework. I will henceforth refer to these <ART+*que*> structures as *semilibre* relative clauses, or *semilibres*, following Brucart (2016). *Semilibres* are the result of an ellipsis process common in Spanish nominal structures, in which nouns can be elided when the syntactic environment contains enough information to make retrieval possible (Brucart, 2016). Whereas in prep-RCs, the article is within the embedded clause as part of the prepositional phrase (10a), *semilibre* RCs are subject or object RCs that modify a partially elided definite antecedent (10b).

- (10) a. [CP [PP P [DP ART \emptyset_{wh}]] que...]
 b. [PP P [DP ART \emptyset_N [CP OP que...]]]

These structures, including selected examples from corpus data, as well as how to differentiate *semilibres* from prep-RCs are discussed in Section 3.1.

2.2.2. Descriptive accounts of article optionality

This section analyses two descriptive sources (Martinez, 1989; Porto Dapena, 1997) that establish parameters for the presence or absence of the article as part of a much broader work that overviews Spanish’s relative pronoun system.

Martinez (1989) sees the article as an extension of *que*, whose role is to prevent homonymy-based ambiguities. In the case of prep-RCs, we can see this in how the article serves to differentiate embedded PPs headed by prepositions *por*, *para*, *desde*, *hasta* from conjunctions like *porque* ‘because’ and *para que* ‘so that,’ and adverbs like *desde que* ‘since’ and *hasta que* ‘until’. He makes the personal observation that use of the article has become more common throughout his lifetime, and the contexts in which it appears to be mandatory seem to have increased. At the same time, the list of prepositions that permit the absence of the article seems to be shrinking. He gives a list of such prepositions: *a* ‘to’, *con* ‘with’, *de* ‘from’, *en* ‘in/on/at’, *sobre* ‘on top’, observing that the list is mostly composed of monosyllabic prepositions. He also notes specifically that the article is mandatory when the prepositional phrase is composed of more than one word, as in (11).

- (11) ... la persona [**gracias a** la que lo conseguiste...]
 the.F.SG person **thanks to** ART.F.SG QUE OBJ.3SG.M achieved.2SG
'...the person thanks to whom you achieved it.'

(Modified from Martinez, 1989, p. 155)

The primary focus of Martinez' account, however, is the impact of syntactic position on the presence of the article. He observes that when the *wh*-phrase has overt case marking, the article is mandatory. This case marker *a* 'ACC.' is polysemous with the preposition *a*, and it exhibits preposition-like properties. It appears before Direct Objects in a complex distribution pattern known as Differential Object Marking (DOM).¹⁰ The article is mandatory when a Direct Object contains DOM, and cannot appear when DOM is absent, i.e., when the *wh*-phrase does not resemble a prepositional phrase (12). Regarding case marking of Indirect Objects, *a* 'DAT.', a phenomenon with a lot of overlap with DOM (Fabregas, 2013), Martinez (1989) claims that the article "might appear or not (although the normal thing is for it to appear)" (Martinez, 1989, p. 154) (13). When the *wh*-phrase is an oblique (i.e., prepositional) argument (14) or an adjunct (15), the article is optional, subject only to the restrictions outlined above.

- (12) Los alumnos a *(los) que suspendió no protestaron
 the.M.PL students ACC ART.M.PL QUE suspended.3SG not protest.3PL
'The students whom (he) suspended did not protest'
- (13) La señora a (la) que se pasó el aviso no asistió
 the.F.SG woman DAT ART.F.SG QUE PSV passed the notice not attended
'The woman to whom the notice was given did not attend'
- (14) No son muchas las personas en (las) que confía
 not be.3PL many the.F.PL persons in ART.F.PL QUE trusted.3SG
'The people in whom he trusted aren't many'
- (15) Ignoró a la persona con (la) que trabajaba
 ignored.3SG ACC the.F.SG person with ART.F.SG QUE worked.3SG
'(He) ignored the person with whom (he) worked'

¹⁰ The study of DOM is one of the most complex and well-researched topics in the syntactic literature of Spanish (for an overview, see: Fabregas, 2013) and well beyond the scope of this research project. For our purposes, it suffices to say that Animacy and Definiteness are the syntactic properties most associated with the presence of this case marker.

(Modified from Martinez, 1989, p. 154-5)

Porto Dapena (1997) does not speculate on the purpose of the article, beyond a brief conjecture that it might have originated through analogy with *semilibre* RCs (1997, p. 28). Although the bulk of his analysis concerns itself with differentiating the article in prep-RCs from the one in *semilibres*, he briefly outlines the following conditions for the <PREP+*que*> variant to appear. In these conditions, the article is truly optional, and its presence/absence is “purely aesthetic” (Porto Dapena, 1997, p. 24). These conditions are that the relative clause must be restrictive (16a), rather than appositive (16b); the antecedent must be **[-human]** (17), and the preposition must be one of the following: *a* ‘to’, *con* ‘with’, *de* ‘from’, *en* ‘in/on/at’, *por* ‘by’.

- (16) a. No tengo herramienta con que arreglar-te el coche
 not have.1SG tool with QUE fix.INF-DAT.2SG the car
‘I don’t have (a) tool with which to fix your car’
- b. *Le regalaron un sombrero, con que está muy contento
 DAT.3SG gifted.3PL a hat with QUE is.3SG very happy.M.SG
 (Porto Dapena, 1997, p. 25)

- (17) La niña a *(la) que le dieron el premio es
 the.F.SG girl DAT ART.F.SG QUE DAT.3SG give.3PL the prize is
 hija de Pepe
 daughter of Pepe
‘The girl to whom the prize was given is Pepe’s daughter’

(Modified from Porto Dapena, 1997, p. 29)

2.2.3. Unique Que hypothesis

The Unique Que hypothesis is most strongly associated with the work of Josep Maria Brucart (1992, 1994, 1999), although Suñer (2001) also links Plann (1980) with this approach. Brucart (1992, 1994) provides a detailed analysis of prep-RCs as part of a bigger project of creating an explanatory theory that can account for the syntactic behaviour of relative pronouns in Spanish.¹¹ He proposes that, for an embedded clause to be relative, it must

¹¹ Specifically, his focus is on pronominal relative pronouns: *que* / PREP+*que* / PREP+ART+*que* (which he considers to be variations of the same linguistic object), *quien* and ART+*cual*.

contain two features: [+QU], which indicates subordination, and [+Rel], which indicates relativization (Brucart, 1992, 1994). He argues that *que*, a complementizer head, contains only a [+QU] feature, and differences in the distribution of relative pronouns stem from how the [+Rel] feature is implemented. In prepositional RCs, [+Rel] is part of a null operator (OP) that pied-pipes the prepositional phrase, including the optional article.

This proposal is supported by evidence that <PREP+ART+*que*> is not a constituent in prep-RCs. The Canary Islands dialect of Spanish contains a superlative construction in which a superlative AdvP is inserted between the article and *que* (Bosque & Brucart, 1991, as cited in Brucart 1994). This superlative (*mas confianza* ‘most trust’ in (18)) is inserted at C’, between the Operator phrase (*con el* OP) at Spec, CP and the Complementizer *que* at the C head.¹²

- (18) El amigo [CP con el OP [C más confianza [C que [TP tengo]]]]
 the.M.SG friend with ART.M.SG most trust QUE have.1SG
‘The friend who I trust the most’ (lit. ‘with whom I have the most trust’)

(Brucart, 1994, p. 463, bracketing mine)

Although it is not central to his analysis, Brucart (1992, 1994) dedicates a fair amount of attention to analysing the article in prep-RCs and the factors licensing its absence. He proposes that the role of the article is to reproduce the φ -features of the antecedent, which makes the presence of the null OP easier to identify. By this, he means both by a potential listener and by the antecedent, acting as a bridge of sorts to ensure that the antecedent-operator relationship remains local.

Brucart (1992, 1994, 1999) identifies three broad constraints on the absence of the article. The first constraint is polarity – the article is only optional when the embedded clause is positive (19) but mandatory when it contains negation (20). The second constraint is definiteness – the article is optional when the antecedent is definite (19), but mandatory when it is indefinite (21). He briefly speculates that, taken together, these two restrictions might indicate that the PP with the article is a referential argument but the one without the

¹² While this exact analysis with multiple C’ levels would not necessarily hold from a modern perspective, it is nonetheless undeniable that *con el* and *que* do not behave like a constituent in this dialect, which does indeed suggest that *que* does not originate low in the tree.

article is not. He suggests that this restriction “might be derived from the intrinsic definiteness of a relative pronoun” (Brucart, 1999, p. 495). That is, when the antecedent is indefinite, it becomes necessary to assert the definiteness of the operator through the article, but this is redundant when the antecedent is definite.

- (19) El dinero de(-l) que disponía ...
 the.M.SG money to-ART.M.SG QUE had.access.3SG
 ‘The money to which he had access...’
- (20) El dinero de-**l** / *de que no disponía ...
 the.M.SG money to-**ART.M.SG** to QUE not had.access.3SG
 ‘The money to which he did not have access...’
- (21) Un dinero de-**l** / *de que disponía ...
 the.M.SG money to-**ART.M.SG** to QUE had.access.3SG
 ‘A money to which he had access...’

(Brucart, 1992, pp.119-20)

The third constraint is better characterized as a series of observations rather than a constraint per-se. Brucart (1999) mentions the syllabicity and homonymy restrictions identified by Martinez (1989), speculating that the effect of syllabicity might be because polysyllabic prepositions, being heavier, interfere with the operator-antecedent relationship in a way that monosyllabic prepositions do not. He also identifies lexical-semantic properties associated with the absence of the article. He observes that antecedents like *modo*, *manera* and *forma* ‘way/manner’ tend to prefer the <PREP+*que*> construction, if not require it. He speculates that sentences with these antecedents might have some kind of predicational properties that differentiate them from standard relative clauses. We return to Brucart’s (1992, 1999) discussion of the factors that enable the absence of the article in Section 5.2.1.

2.2.4. Bare Que hypothesis

The Bare Que hypothesis was first introduced by Rivero (1980, 1982). It proposes that the *que* in prep-RCs is different to the *que* that appears in most types of embedded clauses. While the former is a relative operator (i.e., a *wh*-word), the latter is a complementizer. The main benefit of this approach is that it allows us to make a unified analysis of prep-RCs across Romance languages. Further evidence in favour of this hypothesis comes from the fact that

que (and by extension, ART+*que*) can be substituted by other *wh*-words like ART+*cual* (22b) and *quien* (22c), suggesting that it is a member of that same category.

- (22) a. El empleado con **el que** hablamos
 the.M.SG employee.M with ART.M.SG QUE talked.1PL
- b. El empleado con **el cual** hablamos
 the.M.SG employee.M with ART.M.SG **which** talked.1PL
- c. El empleado con **quien** hablamos
 the.M.SG employee.M with **whom** talked.1PL
'The employee with whom we talked.'

(Rivero 1980, pp. 222-223)

In her response to Brucart (1992, 1999), Suñer (2001) situates the Bare Que hypothesis as the standard analysis of prepositional RCs. She argues that, since the Unique Que hypothesis only accounts for a marginal amount of data, it does not present a strong enough argument to displace the Bare Que hypothesis as the standard analysis of prep-RCs. She also points out that in Spanish, it is not grammatical for a preposition to select a null complement (23a), as evidenced by the ungrammaticality of dangling prepositions in Spanish (23b). Brucart (1992) seems to have anticipated this argument, countering that it is also ungrammatical for a preposition to select a null complement in English (24), and that it is common cross-linguistically for relative operators to be null in contexts that do not otherwise permit the presence of a null argument.

- (23) a. *Anoche hablaron con \emptyset .
 last-night talked.3PL with
- b. *La casa que vivimos en \emptyset durante los ochenta
 the.F.SG house that lived.1PL in during the eighty
'The house that we lived in during the eighties.'

(Suñer, 2001, p. 871)

- (24) *Last night, they talked with \emptyset .

Suñer takes two different approaches to support her argument that the data discussed in Brucart (1992, 1994, 1999) is marginal. The first concerns the superlative construction in

Canarian Spanish, which Brucart (1994) uses to argue that the article is not part of the same constituent as *que*. Suñer points out that this phenomenon would be better analysed as a case of hyperextension, where a pattern of quantifier-raising that is grammatical with superlative *semilibres* (25) is extended to prep-RCs. She backs this argument by pointing out that the phenomenon did not spread to other dialects. However, it is worth mentioning the superlative construction has since been attested in Puerto Rican Spanish (Rohena-Madrado, 2007, as cited in Perpiñan, 2010).

(25) a. son los \emptyset peores que se portan
 are.3PL the.M.PL worst.PL that REFL behave
 '(They) are the ones who behave the worst'

b. son los \emptyset [_{CP} OP₁ [_{C'} [_{QP} peores]₂ [_{C'} que [_{IP} PRO se portan ~~OP₁ peores₂~~]]]]

(Suñer, 2001, p. 870)

The second approach targets the <PREP+ART+*que*> structure itself, which, she argues, is extremely rare in all but the most formal registers. She points to a corpus study of relative clauses in the Spanish of Caracas, Venezuela (Bentivoglio & Sedano, 1993, as cited in Suñer, 2001).¹³ In this corpus, utterances with a <PREP+(ART)+*que*> structure (oblique RCs + direct and indirect object RCs with case marker *a*) only constitute 3.25% of the entire corpus (47/1446). Out of these, only 13/47 contain an article (29.79%). Since <PREP+ART+*que*> utterances constituted less than 1% of all the utterances in the corpus, Suñer argues that “the conditions [for the presence of the article] become practically irrelevant when confronted with actual data” (2001, p. 872). In other words, the fact that the Unique Que hypothesis accounts for article alternation in <PREP+(ART)+*que*> is not a strong argument in its favor, because this structure is not reflective of everyday spoken Spanish.

It is worth mentioning that a corpus by Santana Marrero (2004; see Section 2.4), which centers exclusively on article alternation, shows the same overall trends as Bentivoglio & Sedano, (1993, as cited in Suñer, 2001), with <PREP+ART+*que*> representing approximately a third of prep-RCs, and prep-RCs being relatively rare in their source corpus. So, while it is

¹³ She actually lists several corpora to back this claim, but admits that, of these studies, the Caracas corpus is the only one she considers reliable, as it does not conflate <PREP+ART+*que*> with *semilibres*.

true that the <PREP+ART+*que*> structure is uncommon in Spanish, it seems premature to dismiss its validity as part of the Spanish grammar on those grounds.

2.2.5. Conclusions on the categorization of ‘*que*’

Despite the central importance of the article to the discussion between Bruccart (1992, 1999) and Suñer (2001), I must highlight that the topic of this debate, and the object of analysis for both parties, is that of the categorization of *que*. The presence and optional nature of the article is only a weapon in this debate, rather than the object of analysis in its own right. Nonetheless, the categorization of *que* in prep-RCs is an important structural question that must be addressed before we can delve deeper into the issue of the article.

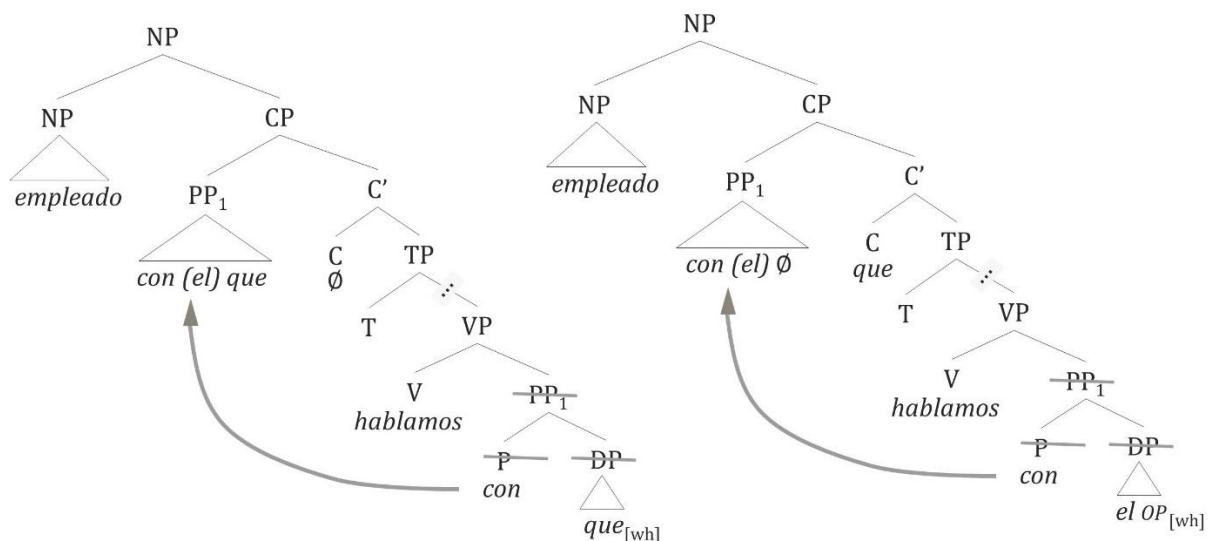


FIGURE 1: Trees for the sentence *El empleado con (el) que hablamos* ‘The employee with whom we talked’ according to the Bare Que hypothesis (left) and the Unique Que hypothesis (right).

Both Bruccart and Suñer present a fairly standard structure for the prep-RC, with a Complementizer and an Operator that undergoes *wh*-movement. As we can see in Figure 1,¹⁴ both proposals are identical in terms of structure, with the only difference being which element is null and which one is overt. For Bruccart, the Complementizer is *que* and the

¹⁴ Both trees presented here represent my best understanding of the hierarchical structures proposed by Bruccart and Suñer. Bruccart (1992, 1994) contains little bracketing and no tree structures; the trees in Suñer (2001) are presented in even less detail than what I have tried to present here.

Operator is null; for Suñer, the Complementizer is null and the Operator is *que*. This is not to say that the issue of categorization is not an important one, but I want to highlight that neither proposal accounts for the article on a structural level.¹⁵

In the case of Suñer (2001), she makes it quite clear that this is because she believes that it is unnecessary to account for the article. To her, the article is only an artifact of prescriptive written language and does not contribute to the syntax of prep-RCs. By contrast, Brucart (1992, 1994, 1999) makes a concerted effort to account for the article, analysing its role and distribution. He clearly identifies its possible function and the kind of syntactic relationships to which it might contribute. Nonetheless, Brucart's proposal does not identify where the article might merge into the tree or what mechanism underlies its role as operator-antecedent liaison.

Of these two options, adopting Brucart's Unique Que Hypothesis is clearly the better choice. This framework acknowledges the article as part of the syntactic structure of prep-RCs and indicates, in very broad terms, where it enters the derivation (i.e., as part of the operator phrase). By contrast, the Bare Que hypothesis either does not acknowledge the article (Rivero 1980, 1982) or actively rejects its belonging in the syntactic structure (Suñer, 2001). Moreover, Brucart (1992, 1994, 1999) provides some promising avenues of exploration and explanation on the topic of article optionality. However, this proposal does not quite provide a satisfactory account of the nature and structural properties of the article. In the next section, I look at whether some of the innovations that have emerged in the last two decades concerning the study of prep-RCs could provide some additional insights.

2.3. Current analyses of prep-RCs

This section examines two broader concepts that underlie the study of prep-RCs, introducing some innovative analyses in the hopes that their insights can be integrated to the study of the Spanish prep-RC. Specifically, I hope that some aspect of these theories can be applied to extend our understanding of the article and article optionality. Section 2.3.1 examines the

¹⁵ It is also worth mentioning that neither account fully accounts for prep-RC data in contexts beyond the scope of this paper. The Bare Que hypothesis cannot account for the superlative construction in Canarian Spanish, while the Unique Que hypothesis might potentially over-generate by not explicitly restricting the possibility that Complementizer *que* could co-occur with an overt operator.

idea of pied-piping and presents Cable's (2010a, 2010b) Q-Theory as an alternative to the traditional feature percolation analysis. Section 2.3.2 gives a brief overview of the study of the relationship between the head and the operator, presenting Cinque's (2020) unifying theory of relative clause structure.

2.3.1. *Pied-piping*

Pied-piping (PiP) refers to a phenomenon where a movement operation associated with a given operator applies to a larger phrase containing this operator (Horvath, 2017). Although most research on feature percolation involves *wh*-phrase movement, PiP can refer to any kind of A-bar movement that is "too large" to be accounted for by a theory's standard story of movement. This means that whether a given movement operation is classified as PiP or not can change depending on both the theoretical approach used and how the relationship between the pied-piping operator and the pied-piped element is defined.¹⁶ Pied-piping of a prepositional phrase is one of the most common types of PiP cross-linguistically, and one of the ones that has been defined most consistently as such (Horvath, 2017).

Feature percolation, a mechanism in which a feature is transmitted beyond the maximal projection of the Head that bears it, has long been the standard solution to the question of pied-piping. Heck (2008) identifies Grimshaw's (1991, as cited in Grimshaw, 2000) theory of extended projections as the most influential percolation-based approach to pied-piping. An extended projection consists of a lexical head (N, A, V) and the functional heads (such as D, P, *v*, etc.) that select it. *Wh*-features can percolate within an extended projection, but not between projections (illustrated in (17) for the relevant pre-movement section of (13a) above, *El empleado con el que hablamos*). Since the preposition is considered part of the extended projection of the NP, the operator can transmit the *wh*-movement feature to P, allowing it to carry the entire prepositional phrase in movement.

¹⁶ For example, before Abney's (1987) DP hypothesis postulated D as the head of the nominal domain, movement of a *which* phrase was characterized as PiP. This is because only heads can trigger movement. Since, at the time, the Determiner *which* was characterized as a Specifier, the fact that it triggered movement of the whole nominal phrase was seen as a type of pied-piping.

(26) [VP hablamos ([PP con [DP el [NP OP]]]]]]]

With the advent of the Minimalist Program (Chomsky, 1995), pied-piping has undergone scrutiny. One of the goals of the Program is to reduce Universal Grammar (UG) to the smallest possible number of mechanisms. Under this principle, the Feature Percolation mechanism should ideally be reduced to a pre-established mechanism like Merge or Agree. This propelled writers like Heck (2008)¹⁷ and Cable (2010a, 2010b) to challenge feature percolation and propose alternative accounts of pied-piping that are compatible with Minimalist principles.

Cable's Q-Theory of pied-piping challenges the concept of pied-piping itself. He argues that PiP is an illusion, caused by the assumption that *wh*-movement is motivated by a lexical feature on the *wh*-word. Instead, he claims that "for [movement] phenomena related to *wh*-operators, the locus of explanation is a distinct element bearing a special semantic & syntactic relationship to the *wh*-operator" (Cable, 2010a, p.1). He calls this element Q. The Q particle takes a phrase containing the *wh*-operator, which can but need not be the *wh*-phrase, as complement. This phrase then moves to Spec, CP motivated by a strong (i.e., EPP-carrying) uninterpretable feature in the C-head.

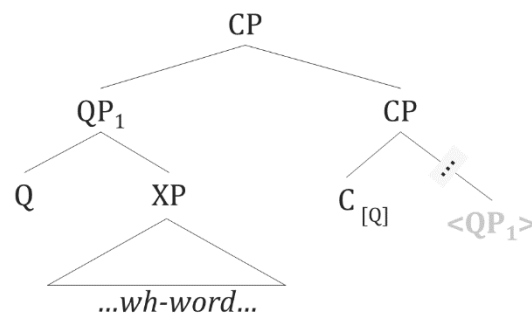


FIGURE 2: The Q-based analysis of *wh*-movement (adapted from Cable, 2010b, p. 567)

Cable's proposal was originally motivated by the study of Tlingit, a Na-Dené language found in Alaska. Tlingit has an interrogative particle, *sá*, that appears at the right edge of a phrase

¹⁷ Heck (2008) uses Optimality Theory to account for pied-piping; it is beyond the scope of this paper.

containing a *wh*-word (27a). Cable argues that this particle is an overt expression of Q (rather than, e.g., an overt C head or a movement marker). Evidence comes from the fact that a *wh*+*sá* string is interpreted as indefinite in its pre-movement position (27b), but interrogative after movement (27a).

(27) a. **Aadóoch** **sá** kwgwatóow yá x'úx' ?
 who.ERG Q will.read this book
 'Who will read this book?'

b. Yá x'úx' akwgwatóow **aadóoch** **sá** .
 this book will.read who.ERG Q
 'People will read this book.'

(Cable, 2010b, p. 569)

In Tlingit PiP structures, *sá* is located directly after the pied-piped phrase (28a); it cannot be contained inside the PiP phrase (28b). To explain this and prevent overgeneration of PiP structures, Cable proposed the QP Intervention Condition constraint, which states that "A QP cannot intervene between a functional head F and a phrase selected by F" (Cable 2010a, p. 57). In (28b), *sá* is ungrammatical because it interferes between the postposition *teen* and the DP.

(28) a. [QP [PP **Aadóo** teen] **sá**] yeegoot ?
 who with Q you.went
 'Who did you go with?'

b. * [PP [QP **Aadóo** **sá**] teen] yeegoot ?
 who Q with you.went

(Cable, 2010a, p.105)

Cable (2010a, 2010b) argues that the Q analysis of Tlingit should be extended to all languages with *wh*-movement. His reasoning is that, if *sá* was a null particle, Tlingit would exhibit the same movement patterns as other *wh*-fronting languages. Therefore, it is worth examining whether these languages might have a null QP layer. Cable identifies three properties that QP-languages share, which can serve as diagnostics for whether a language contains a QP layer. The first property is that possessor *wh*-words cannot be extracted, i.e., possessor structures must be pied-piped (29). The second is that determiner *wh*-words cannot be extracted, i.e., the D head must carry/pied-pipe its complement (30). Finally, *wh*-

words cannot be extracted from a PP, i.e., the preposition/postposition must not be stranded (31). As the examples below demonstrate, Spanish fits all three criteria.

(29) **Cannot extract possessor in Spanish**¹⁸

- a. [DP El señor [CP [DP **cuyo** barco] tú viste ~~cuyo-barco~~]
 the man whose boat you saw.2SG
'The man whose boat you saw'
- b. * [DP El señor [CP **cuyo** tú viste [DP ~~cuyo~~ barco]]
 the man whose you saw.2SG boat

(30) **Cannot extract determiner in Spanish**

- a. [DP **¿Cuál** perro] viste ~~cuál-perro~~ ?
 which dog saw.2SG
'Which dog did you see?'
- b. * **¿Cuál** viste [DP ~~cuál~~ perro] ?
 which saw.2SG dog

(31) **No P-stranding in Spanish**

- a. [PP **¿Con** **quién**] fuiste ~~con-quién~~ ?
 with who went.2SG
'Who did you go with?'
- b. * **¿Quién** fuiste [PP con **quién**] ?
 who went.2SG with

Furthermore, Cable distinguishes between a language like Tlingit and what he calls Limited PiP languages, i.e., those that are more restricted in their *wh*-movement. For example, it is possible to pied-pipe a relative clause in Tlingit (32), but this is ungrammatical in Spanish (33). He proposes that Q in these languages has an unvalued feature that must be valued by a *wh*-word. This subjects Q to the restrictions of an agreement relationship. For example, an unvalued feature cannot probe into a RC, which is an agreement island. Thus, Q cannot enter an agreement relationship with the *wh*-word inside of the RC, crashing the derivation.

¹⁸ This sentence comparison uses a relative clause because the Spanish equivalent to 'whose', *cuyo*, is not grammatical in interrogative clauses in any context.

- (32) [NP [CP **Wáa** kwligeysi xáat] **sá**] i tuwáa sigóo ?
 how it.is.big.REL fish Q do.you.want
 'How big of a fish do you want?'

(Cable, 2010b, p. 572)

- (33) * [CP ¿Un pez que es **qué tan** grande] quieres?
 a fish that is how much big want.2SG

Overall, I find Q-Theory to be a compelling explanation of pied-piping, with clear mechanics and underlying principles. It would be worthwhile to incorporate this theory into a new analysis of prep-RC structure, in order to create a clearer picture of its underlying mechanisms.

2.3.2. Head-Antecedent relationship

As Section 2.2 has shown, the presence/absence of the article in the prep-RC seems to depend at least partially on the properties of the antecedent. Therefore, it is important to examine how the relationship between an antecedent and the relative operator with which it co-refers has been conceptualized in generative history.

The standard analysis in the literature is called the Head-External Analysis (Salzmann, 2018). This is the analysis used in the accounts that I have presented so far, where the antecedent NP or DP takes the relative clause as an adjunct, and a co-indexed operator (*which* in the example) moves to the left edge of the embedded clause (34a). Another popular alternative is the Head-Raising Analysis, according to which the noun that we understand as the antecedent (*book* in 34b) originates inside of the relative clause, moving to the left-most periphery of the *wh*-phrase after the whole phrase has undergone *wh*-movement. The RC is then selected as the complement of a Determiner head (*the* in 34b), deriving the linear order (34b). In the Head-Matching Analysis, the operator is reinterpreted as a full noun that originates in the embedded phrase. As it is phonologically identical to the antecedent, this nominal gets deleted under Identity at spellout (34c).

- (34) a. the book_i [CP [DP which_i]₁ John likes **which₁**]
 b. the [CP [DP book₂ which **book₂**]₁ John likes **which book₁**]
 c. the book_i [CP [DP which ~~book_i~~]₁ John likes **which book₁**]

(Salzmann, 2018, p. 1)

A recent proposal by Cinque (2020) expands upon the Head-Matching analysis to provide a unified account of all types of relative clauses.¹⁹ According to this analysis, all relative clauses can be derived from the same type of double-headed structure. However, I only present those structures that could plausibly represent the prepositional RC in Spanish (i.e., restrictive RCs that have a post-nominal surface position). This structure, represented in Figure 3, contains two identical nominals, henceforth dP. dP is a descriptive label that might refer to a variety of category phrases (NumP, nP, etc.), which crucially must be smaller than DP. The relative clause adjoins as a modifier of the ‘external head’; then one of the heads is deleted through Identity.

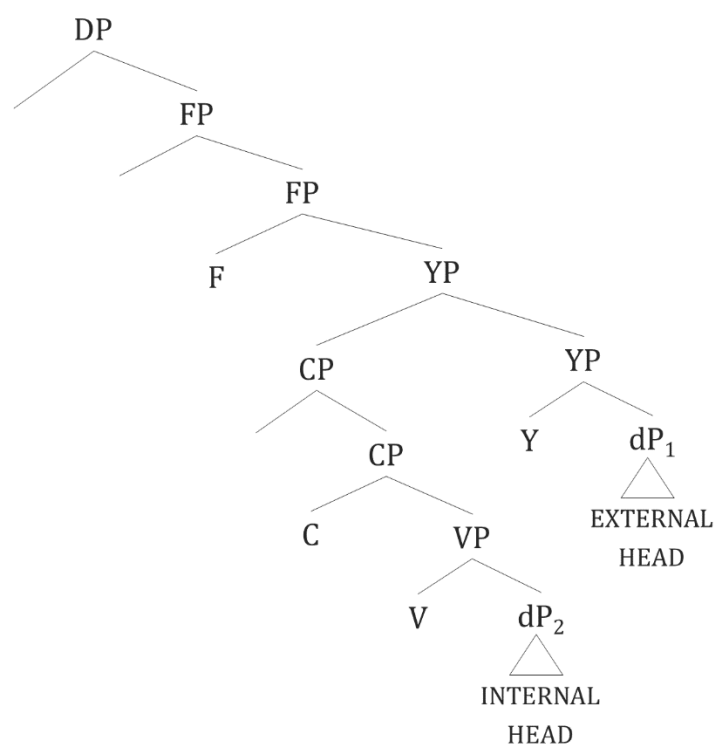


FIGURE 3: Underlying structure of relative clauses (adapted from Cinque, 2020, p. 15).

¹⁹ Cinque (2020) is strongly based on Kayne’s (1994) anti-symmetry framework, which further constrains the standard minimalist framework through some additional assumptions. For the purposes of understanding Cinque’s proposal, what we need to know is any unit that is merging into the derivation is assumed to do so on the left side of the current phrase; this occurs across all languages and various surface orders are derived through cyclical movement. For the purposes of this framework, c-command relationships must be asymmetrical – that is, rather than two sisters c-commanding each other, “X asymmetrically c-commands Y iff X c-commands Y and Y does not c-command X” (Kayne 1994, p. 15).

Cinque (2020) is best understood as a Head-Matching analysis, although it incorporates aspects of the Head-Raising analysis, in particular the idea that the overt head might be the nominal inside of the RC. The RC-internal nominal raises to Spec, CP during *wh*-movement. As it is higher in the tree than the external head, it licenses the deletion of the dP_1 (35a). However, if the dP_1 moves up the nominal tree, then it licenses the deletion of dP_2 instead (35b).

- (35) a. [DP the [YP [CP [dP_2 two nice books]₁ that John wrote [~~dP_2 -two nice books~~]₁] [YP Y [~~dP_1 -two nice books~~]]]]
 b. [DP the [dP_1 two nice books]₂ [YP [CP [~~dP_2 -two nice books~~]₁ that John wrote [~~dP_2 -two nice books~~]₁] [YP Y [~~dP_1 -two nice books~~]₂]]]]

(adapted from Cinque, 2020, pp. 17-18)

Whether the overt head in any given relative clause is internal or external can be diagnosed through a variety of *c*-command and scope tests, all of which are beyond the scope of this thesis. For the Spanish prep-RC, it suffices to look at the linear order. As (36a) below shows, an externally-headed structure (one where the operator phrase is deleted) results in the linear order of a <PREP+*que*> sentence. An internally-headed structure (where the antecedent is deleted), by contrast, results in an ungrammatical linear order where the preposition is perceived to be “outside” of the relative clause (36b).²⁰

- (36) a. la mesa [CP [PP en ~~la mesa~~]₁ que puse el libro **en la mesa**₁]
 b. * ~~la mesa~~ [CP [PP en la mesa]₁ que puse el libro **en la mesa**₁]

Given how straight-forwardly Cinque’s (2020) double-headed structure can account for the derivation of <PREP+*que*> sentences, this framework seems like a promising avenue for exploring the Spanish prep-RC and the article alternation within it.²¹ I further investigate this possibility in Chapter 5. In order to set the stage for the experimental components of this

²⁰ Of course, it would still be possible to derive the linear order in (27b) by arguing that the internal head moves to Spec,P. In fact, this is what Kato & Nunes (2009) propose for prep-RCs in Brazilian Portuguese, a language with the same pied-piping pattern as Spanish. However, the authors do not motivate this movement beyond the need to derive the linear word order (following Kayne’s (1994) antisymmetry framework). Since this approach does not offer any obvious advantages, I see no reason to pursue it.

²¹ Another benefit of Cinque (2020) is its compatibility with Brucart’s (1992) Unique Que Hypothesis. Cinque briefly touches upon the data in Brucart (1992), broadly contrasting the Canarian Spanish data with comparable structures in other languages (2020, pp. 59-62).

research project, let us first look at the data that has been gathered on article optionality and whether it matches the descriptive accounts and theoretical understanding of this topic.

2.4. Empirical study of Article Optionality in prep-RCs

As Section 2.2 shows, formal research into article optionality in prep-RCs seems to be at an impasse. Authors like Martínez (1989) and Brucart (1999) commented that the article seemed to be in the process of a change of the grammar, becoming mandatory in contexts where it was previously optional. As these claims were made as early as 30 years ago, it is important to keep observing the behaviour of this structure and re-evaluating the claims. In particular, Suñer's (2001) assertion that the <PREP+(ART)+*que*> structure is not a part of "everyday Spanish" is one that must be corroborated empirically.

In this section, I evaluate Santana Marrero (2004), a synchronic²² corpus analysis that compiles and examines article optionality in prep-RCs across multiple Spanish dialects. This paper is unique in that it focuses exclusively on article optionality. While this paper is far from the only corpus to study Spanish prep-RCs, other corpora (e.g, Bentivoglio & Sedano, 1993, as cited in Suñer, 2001), only include prepositional RCs as a small section of a larger research project. Section 2.4.1 presents the methodology used and claims made in Santana Marrero (2004), and Section 2.4.2 gives a critique of the paper.

2.4.1. Summary of Santana Marrero (2004)

The data in Santana Marrero (2004) were extracted from an oral corpus of interviews of "educated class" participants in 12 cities across the Hispanic world (Samper et al. 1998, as cited in Santana Marrero, 2004).²³ Out of the original data (168 interviews giving 84 hours of recordings), the researchers extracted all prepositional RCs, then excluded an unreported number of <PREP+ART+*que*> strings that did not allow for a no-article variant (Santana Marrero, p. 69). These strings were headed by the prepositions *a* and *con*, and were

²² Although Santana Marrero (2004) is the only synchronic study, there has been significant research into the emergence of the <PREP+ART+*que*> variant from a diachronic perspective. This body of work (Blas Arroyo & Vellón Lahoz, 2017, 2018 and sources within) indicates that the presence of the variant with the article has seen a slow but steady increase in popularity since the XVIII century, while still not becoming the dominant form by the mid-XX century.

²³ *Macrocorpus de la norma lingüística culta de las principales ciudades del mundo hispánico* 'Macrocorpus of the educated linguistic norms of the main cities in the Hispanic world.'

characterized by having an animate antecedent.²⁴ Truncated sentences were also excluded. These resulted in a total of 527 prepositional RCs, whose distribution is reported in Table 1.

		<PREP+ART+ <i>que</i> >		<PREP+ <i>que</i> >		Total
<i>a</i>	'to, ACC'	20	68.97%	9	31.03%	29
<i>con</i>	'with'	20	46.51%	23	53.49%	43
<i>de</i>	'of, from'	17	70.83%	7	29.17%	24
<i>en</i>	'in, on, at'	101	24.10%	318	75.89%	419
<i>por</i>	'for'	11	91.67%	1	8.33%	12
Total		169	32.07%	358	67.93%	527

TABLE 1: Reproduction of Table 1 in Santana Marrero (2004, p. 73), showing the distribution of prep-RCs in the corpus sorted by preposition.

Santana Marrero coded these utterances for the following factors and analysed how each factor affects the distribution of the article for each preposition. The factors examined were: antecedent definiteness, number and animacy; sentence polarity; semantic type of the prep-RC (restrictive vs. appositive); the presence of an identical <PREP+ART> string in the matrix clause (e.g., (37)); and whether there was distance between the antecedent and the RC. In addition, <*en*+(ART)+*que*> utterances were coded for the semantic character of the antecedent – specifically, identifying antecedents of time, place and manner. Syntactic position was coded for <*a*+(ART)+*que*> utterances, following the observations of Martínez (1989).

- (37) ...los **de la** segunda generación **de (la)** que estábamos
 ART.M.PL **of THE.F.SG** second generation **of ART.F.SG** QUE were.1PL
 hablando...
 talking
 ‘...those of the second generation of which we were talking...’

(Santana Marrero 2004, p. 80, emphasis mine)

²⁴ It is worth mentioning here that the <*a+que*> variant with an animate antecedent was not attested in the historical data (Vellón Lahoz, 2020); instead, <*a+ART+que*> alternated with <*a+quien*>.

Overall, the corpus showed a preference for the absence of the article, with 67.93% (358/527) of tokens having no article. The prepositions *a*, *de*, and *por* had a preference for <PREP+ART+*que*>; *con* and *en* had a preference for <PREP+*que*>. Of the factors examined, they found that, for most prepositions, the absence of the article conformed to the factors that had been described in the literature – antecedents were definite, singular and inanimate, and the embedded clause had positive polarity. They found either no pattern or too few examples of factors like repetition, distance, and semantic type. However, utterances with *en* did not conform to these patterns. That is, <*en+que*> was the more common variant even in sentences with an indefinite antecedent and/or negative polarity (38).

- (38) Había que hacer un viaje *en una época en que no*
 had to make.INF a trip at a.F.SG time in QUE not
habían cauchos.
 be.EXT rubber
 '(One) had to make a trip at a time where rubber did not exist.'

(Santana Marrero, 2004, p. 82)

Santana Marrero (2004) concludes that it is becoming more prevalent to drop the article across Spanish dialects, particularly in Latin American cities. She observes that the factors that restrict the <PREP+*que*> form do not apply across the board, and exceptions were particularly common with (but not unique to) the preposition *en* (Santana Marrero 2004, p. 88). This increased prevalence of <PREP+*que*>, and the absence of mandatory constraints for monosyllabic prepositions, led Santana Marrero (2004) to suggest that we might be seeing a change in progress towards the disappearance of the article.

2.4.2. Critique of Santana Marrero (2004)

Although Santana Marrero (2004) must be commended for the breadth and focus of its approach, the paper contains some major methodological and analytical flaws that cast doubt on the validity of its conclusions. The most salient issue is the absence of in-depth statistical analysis. Santana Marrero (2004) does not use chi-square analysis or any other type of statistical testing to determine significance. Counts and proportions are reported as 'preferences' and 'tendencies,' but have no predictive power. Therefore, the conclusions in

this paper can only be understood as general observations, rather than as statistical predictions.

The fact that some *<a+ART+que>* and *<con+ART+que>* tokens were removed for not being truly optional is also a concern, as it misrepresents the distribution of the article with these prepositions. It is also unclear how many tokens were removed, or what criteria were used to determine removal. In the case of *a*, this is particularly concerning, since the properties that make the article mandatory are likely related to animacy and syntactic type, two factors that the author identified as being of interest and coded.

Moreover, while the distribution of each factor was reported per preposition, the conclusions about these factors generalize across all prepositions. This is particularly concerning due to the disproportionate representation of *<en+(ART)+que>* tokens, which account for 79.51% (419/527) of the corpus. As Santana Marrero acknowledges, this preposition exhibits a different pattern of behaviour from the other four. Nonetheless, overall proportions are reported as if *<PREP+que>* was dominant across the board, not acknowledging that the behaviour of *en* might be skewing the results. This is compounded by the lack of statistical analysis, which might have accounted for this difference in distribution.

Although Santana Marrero (2004) provides some interesting initial evidence of the contexts in which *<PREP+que>* is available, it is not methodologically reliable. As the optionality of the article has been identified as a potential change in progress, it would also be useful to have more recent data. The following chapter presents the exploratory phase of this research project, in which I conduct my own corpus analysis on the factors that influence article optionality in prepositional relative clauses.

3. CORPUS ANALYSIS

As discussed in the previous chapter, research into article optionality in prep-RCs has stalled and would greatly benefit from empirical data. Although Santana Marrero (2004) is an attempt to fill this gap in the literature, the study has some issues in terms of its methodological and statistical validity, see Section 2.4.2. For these reasons, my research begins by conducting a corpus study that provides more current information on the usage patterns of prep-RCs and a more reliable statistical analysis.

Section 3.1 presents how the prep-RCs utterances that compose my corpus were identified and extracted. Section 3.2 presents the coding scheme. Section 3.3 presents the results and Section 3.4 discusses the implications of these results and how they shape the next steps of this research project.

3.1. Methodology

This corpus is based on a free sample from the *Corpus del Español* ('Corpus of Spanish', Davies, 2004), which compiles and tags web sources such as blogs, forum posts and news articles.²⁵ The free sample contains two million words' worth of randomly selected utterances from the corpus and has been stripped of metadata (source and nation of origin). I ran a script on Python using the Natural Language Toolkit (Bird, Loper, & Klein, 2016), which identified and extracted sentences that contained the following string: one of this set of monosyllabic prepositions (39a) + (optionally) one of the following definite articles (39b) + *que*.²⁶

- (39) a. *a* 'to, ACC', *con* 'with', *de* 'of, from', *en* 'in/on/at, about', *por* 'by', *sin* 'without'
 b. *el, la, los, las* 'the'

This set of prepositions was chosen in accordance with the literature, as they have been identified in most of the previous work as being capable of licensing the absence of the article (see Section 2.2). In addition to the five prepositions identified in the literature (*a, con, de, en*

²⁵ In a corpus, it is ideal to use language that is most representative of spontaneous speech, as written corpora tend to be edited and therefore overrepresent prescriptive forms. An advantage of using this particular corpus is that it uses a mixture of edited and unedited sources.

²⁶ The corpus search was also instructed to account for the cliticised strings *al* 'to the.M.SG' and *del* 'of the.M.SG'. This turned out to be unnecessary, as Corpus del Español separates all commonly used clitics into their components.

and *por*), I included the monosyllabic preposition *sin* ‘without’. This was because *sin*, while being monosyllabic (a condition predicted to permit <PREP+*que*>), it is also a negative polarity item (predicted to restrict <PREP+*que*>). Due to the large amounts of non-target prepositional phrases in the original results, I wrote a new script that excluded common false positives, e.g., *después de* ‘after’, *antes de* ‘before’, *pese a* ‘despite’, etc.

This new script gave out 4400 utterances containing the strings <preposition+*que*> or <preposition+article+*que*>, of which a set of 1000 was randomly selected to form the body of this corpus. This set was then cleaned up by splitting utterances that contained more than one <preposition+(article)+*que*> string and removing sentence fragments that cut off the embedded clause (sentence fragments that cut off the matrix clause were only removed if the antecedent was partially or totally cut off). Following Santana Marrero (2004), I also removed instances of the expression *en la medida en que* ‘insofar as.’ After clean-up, the total number of utterances was 1028.

Then I set to tagging and separating strings with the <PREP+(ART)+*que*> surface form that were not prep-RCs. Of the 559 utterances removed, 421 were <PREP+*que*> strings and 138 were <PREP+ART+*que*> strings. False positive <PREP+*que*> strings mostly consisted of subject and object embedded clauses that were selected by a prepositional phrase. Unlike English, these clauses are not usually headed by a null complementizer in Spanish. These were quite straightforward to identify, as the string is preceded by a verb or noun that is known to select a PP, and no potential antecedent can be identified. For example,²⁷ in (40), the verb *olvidar* ‘forget’ selects a prepositional phrase headed by *de* ‘about’. This preposition selects a CP as complement, creating a <*de+que*> string.

(40) El neoliberalismo pretende que la gente se **olvide** **de** [que
the neoliberalism intend.3SG that the people CL forget.SJV of that
tiene derechos]
has.3SG rights

‘Neoliberalism intends for people to forget that they have rights.’

False positive <PREP+ART+*que*> strings consisted of *semilibre* relative clauses. As discussed in Section 2.2.1, *semilibres* are subject or object RCs that have been selected by a partially elided

²⁷ Unless otherwise indicated, all example sentences in this chapter are sourced from the corpus.

antecedent (Brucart, 2016). In the case of these false positive tokens, the article-like antecedent is selected by a preposition. The structural difference between prep-RCs and *semilibre* RCs is best illustrated through an example. Sentence (41a) is a prepositional relative clause.²⁸ As illustrated in (41b), the article *los* is within the prepositional phrase *de los*, which originated as an argument of *destacaría* ‘I would highlight’ and remains inside of the CP after movement. By contrast, (42a) is a *semilibre*. As we can see in (42b), the prepositional phrase and the article originate outside of the embedded clause, serving as its antecedent. As we can see, the NP that might initially look like an antecedent, *una novela muy buena* ‘a very good novel’, does not match in ϕ -features with the article, and the verb *enganchar* ‘to hook in’ does not select PPs headed by *de*.

- (41) a. Por mis manos han pasado unos 35 libros,
 through my hands have passed a.M.PL books
de los que destacaría estos
 of ART.M.PL QUE highlight.COND.1SG PROX.M.PL
‘Through my hands have passed some 35 books, of which I would highlight these (ones)’
- b. [DP ...35 libros [CP [PP de [DP los \emptyset_{wh}]] que...]]
- (42) a. es una novela muy buena, **de las que** realmente engancha
 is.3SG a.F.SG novel very good of ART.F.SG QUE really hooks-in
‘(It) is a very good novel, of the (ones) that really hooks (you) in’
- b. [PP de [DP las \emptyset [CP OP que...]]]

As the example above shows, it can be tricky to differentiate between a prep-RC and a *semilibre* RC, particularly in cases where the antecedent of a prep-RC is null or not immediately evident, or where the meaning of the RC is ambiguous or hard to parse. In the example pair shown above, both embedded clauses have similar meanings and topics; the mismatches that I identify for (32a) might just be non-standard dialectal features or speech errors. I used a diagnostic test to differentiate between these two types of RCs and applied

²⁸ This sentence is a shortened version of the original that was found in *Corpus del Español*. The full sentence is: “**Por mis manos han pasado unos 35 libros** de todos los gustos y calidades **de los que destacaría** *La familia de Pascual Duarte* de Camilo José Cela, *El hombre que fue Jueves* de Chesterton, *El tercer hombre* de Graham Greene o *Territorio Comanche* de Pérez Reverte.”

it to every <PREP+ART+*que*> string whose structure was not immediately evident. The diagnostic consists of two substitution tests that replace part of the string with a structurally identical synonym with the same φ -features. To test for a prep-RC, I substitute the relative pronoun <ART+*que*> for <ART+*cual*> ‘which’ (Suñer, 2001).²⁹ To test for a *semilibre* RC, I substitute the antecedent for the distal demonstrative pronoun *aquel(/la/los/las)* ‘that/those (ones)’ (Porto Dapena, 1997).

By applying these diagnostics to the two utterances above (41-42), the structural differences between them are highlighted. For the prep-RC (41), substituting <*los+que*> for <*los+cuales*> does not create a meaning change, and the sentence stays grammatical (43a). However, substituting <*los+que*> for <*aquellos+que*> creates a nonsensical sentence, as the referent of *aquellos* cannot be found through the linguistic context (43b). By contrast, substituting <*las+que*> for <*aquellas+que*> makes the meaning of the *semilibre* utterance (42) clearer. With *aquellas*, the fact that there is an external referent (the set of novels that hook you in) becomes more evident (44b). However, substituting <*las+que*> for <*las+cuales*> attempts to force an interpretation that is not available (44a). We would have to reconstruct the PP as an argument of *enganchar*, which is ungrammatical.³⁰

- (43) a. Por mis manos han pasado unos 35 libros,
 through my hands have passed a.M.PL books
de los cuales destacaría estos
 of ART.M.PL which.PL highlight.COND.1SG PROX.M.PL
 ‘Through my hands have passed some 35 books, of which I would highlight these (ones)’
- b. *Por mis manos han pasado unos 35 libros,
 through my hands have passed a.M.PL books
de aquellos que destacaría estos
 of DIST.M.PL that highlight.COND.1SG PROX.M.PL

²⁹ As discussed earlier, “relative pronoun” is an umbrella term that refers to complementizer + operator, of which one is often null. The *que* → *cual* substitution substitutes these two elements. That is, [\emptyset_{wh} *que*] is substituted for [*cual* \emptyset_c].

³⁰ That said, I must acknowledge that (34b) (and indeed, (32) above) is technically ungrammatical. We would normatively expect the verb *enganchar* ‘to hook in’ to express number agreement with the subject *las que* (32) / *aquellas que* (34b), that is, to be *enganchan* instead of *engancha*. Nevertheless, this error is easy to identify as such and does not impede interpretation.

- (44) a. *es una novela muy buena, **de las cuales** realmente
 is.3SG a.F.SG novel very good of ART.F.SG which.PL really
 engancha
 hooks.in
- b. es una novela muy buena, **de aquellas que** realmente
 is.3SG a.F.SG novel very good of DIST.F.PL QUE really
 engancha
 hooks.in

'It is a very good novel, (one) of those that really hooks (you) in'

After separating out these non-target utterances, there were 469 prep-RCs left. In the following section, I present how these utterances were coded.

3.2. Coding and Classifying

Each utterance was tagged for the following phrasal factors: (i) Sentence Type, which tagged whether the article was present or absent; (ii) Preposition; (iii) Article; (iv) Finiteness; and (v) Polarity.³¹

The Preposition and Article columns just kept track, respectively, of the head of the pied-piped PP and the form of the article within this PP. Finiteness refers to whether the embedded verb is inflected for a finite tense. Non-finite verbs have been identified (Plann, 1980; Brucart, 1992; Porto Dapena, 1997) as potentially correlating with <PREP+*que*> in contexts where the prep-RC is preceded by a null antecedent. Polarity, i.e., whether the embedded clause has been negated or not was identified by Brucart (1992) as conditioning the presence of the article (see Section 2.2.3).

In addition to these structural factors, I coded the syntactic properties of the antecedent that the literature has identified as conditioning the presence or absence of the article. These properties are: (i) Definiteness, (ii) Number, (iii) Gender, and (iv) Animacy. Brucart (1992)

³¹ Although Porto Dapena (1997) identifies semantic type (i.e., whether the RC is restrictive or appositive) as a factor that influences the presence or absence of the article, I do not code for semantic type in this corpus. As discussed in Section 2.1.2, semantic type cannot be reliably diagnosed through spelling. However, it is worth mentioning that, from a quick visual inspection, I was only able to identify five <PREP+*que*> utterances that could conceivably be interpreted as appositives, which lends support to Porto Dapena's assertion.

asserted that <PREP+*que*> sentences are only grammatical when the antecedent is definite, although Santana Marrero (2004) found that it was common for utterances with an indefinite antecedent to have no article when the preposition was *en*. Grammatical number was first examined in Santana Marrero (2004), on the basis that, since the article and antecedent are connected through agreement, ϕ features might impact the presence of the article. Following this logic, I included the factor of Grammatical Gender.³² Porto Dapena (1997) proposed that only [-human] antecedents permit the absence of the article in prep-RCs, which I have interpreted to refer to grammatical animacy. While the other antecedent properties are binary, this factor has three levels: Animate, Inanimate, and Ambiguous. The Ambiguous tag was used when a normally inanimate antecedent was used in a way that is usually reserved for animate entities, i.e., was personified. For example, in (45), the antecedent *el banco* ‘the bank’ is referring to an inanimate organization that acts as Agent of the matrix verb *analizar* ‘to analyze’.

(45) El banco a el que solicites una hipoteca
 the bank to ART.M.SG. QUE request.SJV.2SG a mortgage
 analizará el reporte.
 examine.FUT.3SG the report

‘The bank from which you request a mortgage will examine the report.’

To examine the structural relationships that might have been overlooked or only partially examined in Santana Marrero (2004), I coded for the theta role of the <PREP+(ART)+*que*> prepositional phrase within the embedded clause. I was particularly intrigued by Santana Marrero’s (2004) assertion that <*en+que*> was predominant with antecedents that express time and manner. This seems to hint at an adjunct/argument distinction, and I was interested to see whether this distinction might extend beyond the one preposition. Although I also coded the theta role of the antecedent for completeness’ sake, this factor was not included in my analysis. The list of theta roles tagged is presented below.

- ACTOR: Non-sentient being that undertakes an action. Only relevant for antecedents.

³² Another advantage of including Gender is that, since cliticization only occurs in M.SG contexts (*a+el = al, de+el = del*), this will allow us to track indirectly whether being in a cliticized construction affects the presence of the article.

- AGENT: Sentient being that intentionally undertakes an action. Only relevant for antecedents.

- BENEFICIARY: Entity that positively or negatively benefits by an action not performed by them.

(46) ...los cuatro o cinco a los que hay que ganarles.
 the.M.PL four or five to ART.M.PL QUE have to defeat.OBJ.3PL
'...the four or five (competitors) that (we) have to win against.'

- CAUSER: Entity that causes an action.

(47) ...la dictadura política por la que tanto drama hace.
 the.F.SG dictatorship political over ART.F.SG QUE so.much drama make.3SG
'...the political dictatorship over which (he) throws such a fit.' (lit. makes so much drama)

- COMMITATIVE: Used to tag *con* utterances almost exclusively. Identifies a partner in an action.

(48) ...recomendar a esos profesionales con los que
 recommend.INF ACC those professionals with ART.M.PL QUE
 ha trabajado
 have.2SG work.PRT
'...to recommend those professionals with whom you have worked...'

- EXPERIENCER: Sentient being undergoing an experience (emotions, sensory feedback, etc.)

(49) ...las emociones que sentimos todos a los que nos
 the emotions that feel.1PL all.M.PL A ART.M.PL QUE CL.2PL
 gusta correr
 enjoy running
'...the emotions that all (of us) who enjoy running felt'

- GOAL: Place where an action ends.

(50) Hay muchos bares y clubes gays a los que podés ir.
 BE.EXT many bars and clubs gay to ART.M.PL QUE could.2PL go
'There are many gay bars and clubs to which you could go.'

- INSTRUMENT: Entity used to facilitate an action.

(51) ...la carta con la que Virginia Woolf se despidió de
 the.F.SG letter with ART.F.SG QUE Virginia Woolf said.goodbye of
 su marido
 her husband

'...the letter with which Virginia Woolf said goodbye to her husband'

- LOCATION: Place (literal or metaphorical) where an action takes place or where an entity is situated.

(52) Hay un poema de Khalil Gibran en el que habla de
 BE.EXT a.M.SG poem by Khalil Gibran in ART.M.SG QUE talks of
 la tristeza
 the sadness

'There is a poem by Khalil Gibran in which he talks about sadness.'

- POSSESSOR: Used to tag almost exclusively *de* utterances (with the exception of a single utterance of *pertenecer a* 'belong to') that indicate a possession relationship. Indirectly, keeps track of *de* in its role as genitive case marker.

(53) Una buena pregunta de la que no tengo respuesta.
 a.f.sg good question of ART.F.SG QUE not have.1SG answer

'A good question whose answer I do not have.'

- SOURCE: Place where an action starts.

(54) ...un huevo a el que se le ha quitado su núcleo
 an.M.SG egg from ART.M.SG QUE PSV CL has remove.PRT its nucleus

'...an egg from which its nucleus has been removed'

- STIMULUS: Entity that causes an experience to occur.

(55) voy a encontrar un trabajo creativo por el que
 go.1SG to find a.M.SG job creative for ART.M.SG QUE
 me apasionaré
 CL.1SG get-passionate.FUT.1SG

'I am going to find a creative job for which I will become passionate'

- THEME: Entity to which an action is done or that is undergoing movement; "elsewhere."

(56) ...era la seña a que usted se refiere.
 was the.F.SG sign to QUE you OBJ.3SG refer.2SG

'...(it) was the sign to which you were referring.'

- TIME: Temporal LOCATION.

(57) ...hubo épocas en que esas aguas conservaban su
 BE.EXT.PST times in QUE those waters maintained their
 limpieza natural.
 cleanliness natural

'...there were times in which those waters kept their natural cleanliness.'

In addition to the theta roles above, the following tags are not generally considered to be theta roles, but were used to track structural distinctions in cases where the PP or antecedent did not appear to have a theta role.

- MANNER: Used to tag PPs that provide additional information about how an action was undertaken.

(58) Es comprensible por tanto el secreto con que son guardadas
 is understandable therefore the.M.SG secrecy with QUE are.3PL kept
'(It) is understandable, therefore, the secrecy which with they are kept.'

- STATE: Indicates the predicate in a copula or otherwise predicational sentence. In example (59), antecedent *figuras* 'figures' is the predicate of the verb *convertirse* 'to become' and [_{PP} *a las que*] is the predicate of *ser* 'be'.

(59) ...para convertirse en figuras a las que aspiramos ser
 to turn.INF.REFL into figures to ART.F.PL QUE aspire.1PL be.INF
 algún día
 someday

'...(in order) to become role models (lit. figures) to which we aspire to be (like) someday'

- Unclassifiable: Used to tag antecedents that were fragments or null. In example (60), the utterance does not seem to have any verb or preposition that could have assigned a theta role to antecedent *lugar* 'place.'

(60) ...designado regidor de El Cacao, **lugar** en el que residía...
 appointed.PRT ruler of El Cacao place in ART.M.SG QUE resided.3SG
'...appointed ruler of El Cacao, (the) place in which he resided...'

My predictions for these factors are: For the categories presented here that were originally examined in Santana Marrero (2004), I predict that the analysis of this corpus will support her results. That is, we will find that <PREP+que> favours antecedents that are definite,

singular and inanimate. Beyond what has been previously established, I predict that preposition will have a significant effect on the presence/absence of the article. As this work is exploratory in nature, I do not make specific predictions for the categories I have introduced, i.e., Finiteness, Gender, and Theta Role.

3.3. Statistical report

The statistical analysis was done in R (R Core Team, 2021). It consisted of chi-square tests where Sentence Type was always treated as one of the variables. Chi-squared analyses indicate whether two categorical variables are independent of each other, that is, whether the distribution of values in variable A (in our case, the presence or absence of the article) is caused by the distribution of values in variable B. Initial analyses gave out error messages, indicating that one or more levels had too few members to conduct a reliable analysis. I removed two utterances whose prepositions only showed up once: one instance of *sin* ‘without’ and one instance of *frente a* ‘in front of’.

I also removed 10 utterances that contained a null antecedent, as they were throwing off the analysis of antecedent properties.³³ These utterances consist of sentences that can only be interpreted if they contain a null antecedent, and this null antecedent is easily retrievable from the common ground. For example, in (51) the antecedent refers to resources used for self-defence. Due to the small number of these utterances, they have been deemed beyond the scope of this investigation. Nonetheless, I must highlight their unique properties, as they might be a fruitful source for future research into article optionality. Null-antecedent prep-RCs were evenly distributed among the prepositions (2 *a*; 3 *con*, 2 *de*, 2 *en*, 1 *por*). The majority of these constructions (7/10) had a null antecedent, and half were part of an infinitive phrase (all of which were <PREP+*que*> utterances).

(61) Como nunca antes, tenemos [DP Ø [CP con que salir a defendernos]]
 like never before have.1PL with QUE go-out to defend.REFL.1PL
 ‘Like never before, we have (the resources) with which to go defend ourselves’

Having done that final culling, I analysed the remaining 457 utterances. Of these, 345 contained an article inside the prepositional phrase (75.49%), and 112 did not (24.51%).

³³ Since finiteness was mainly relevant to these constructions, Finiteness was not considered as an independent variable in the final analysis.

Chi-square analyses found significant effects for the antecedent factors of Definiteness, Number and Animacy, but not for Gender ($\chi^2 (1) = 0.635, p = 0.426$). That is, <PREP+*que*> sentences were significantly more frequent when the antecedent was definite (97/290, 33.45%) rather than indefinite (15/167, 8.98%) ($\chi^2 (1) = 32.977, p < 0.001$); ; singular (87/318, 27.36%) rather than plural (25/139, 17.99%) ($\chi^2 (1) = 4.1001, p = 0.043$); and inanimate (110/400, 27.50%) rather than animate (1/44, 2.27%) or ambiguous (1/13, 7.69%) ($\chi^2 (2) = 15.68, p = 0.0004$). <PREP+*que*> sentences were also significantly more frequent in cases where the embedded sentence had positive polarity (111/429, 25.87%) rather than negative (1/28, 3.57%) ($\chi^2 (1) = 5.9125, p = 0.015$).

	Definite	Indefinite	TOTAL
<PREP+ART+ <i>que</i> >	193	152	345
<PREP+ <i>que</i> >	97	15	112
TOTAL	290	167	457

TABLE 2: Contingency table for the distribution of article presence/absence according to the definiteness of the antecedent.

	Singular	Plural	TOTAL
<PREP+ART+ <i>que</i> >	231	114	345
<PREP+ <i>que</i> >	87	25	112
TOTAL	318	139	457

TABLE 3: Contingency table for the distribution of article presence/absence according to the grammatical number of the antecedent.

	Inanimate	Animate	Ambiguous	TOTAL
<PREP+ART+ <i>que</i> >	290	43	12	345
<PREP+ <i>que</i> >	110	1	1	112
TOTAL	400	44	13	457

TABLE 4: Contingency table for the distribution of article presence/absence according to the animacy of the antecedent.

	Positive	Negative	TOTAL
<PREP+ART+ <i>que</i> >	318	27	345
<PREP+ <i>que</i> >	111	1	112
TOTAL	429	28	457

TABLE 5: Contingency table for the distribution of article presence/absence according to the polarity of the embedded clause.

I also found that the preposition had a significant effect on whether the article is present or absent ($\chi^2 (4) = 55.247$, $p < 0.001$). As Table 6 shows, utterances with the preposition *en* constitute more than half of the corpus, and 81.25% of utterances without an article. My next step was to find interactions between Preposition and the other significant factors. Due to the relatively small number of no-article utterances found for the other prepositions, it was only possible to reliably conduct these tests for the *<en+(ART)+que>* subset.

Preposition	Tokens	Article	No article	Proportion with no article	Proportion of total 'no article' tokens	Proportion of total tokens
EN	245	154	91	37.14%	81.25%	53.61%
CON	51	38	13	25.49%	11.61%	11.16%
A	93	86	7	7.53%	6.25%	20.35%
POR	34	33	1	2.94%	0.89%	7.44%
DE	34	34	0	0.00%	0.00%	7.44%
Total	457	345	112	24.51%	100%	100%

TABLE 6: Distribution of article per preposition

For the *en* subset, utterances without the article are significantly more frequent when the antecedent is definite (77/141, 54.61%) rather than indefinite (14/104, 13.46%). They are also more frequent when the antecedent is singular (72/178, 40.45%) rather than plural (19/67, 28.36%), but not significantly so ($\chi^2 (1) = 2.552$, $p = 0.11$). Chi-square tests for Animacy ($\chi^2 (2) = 3.016$, $p = 0.221$) and Polarity ($\chi^2 (1) = 2.726$, $p = 0.099$) gave unreliability warnings.

	Definite	Indefinite	TOTAL
<en+ART+que>	64	90	154
<en+que>	77	14	91
TOTAL	141	104	245

TABLE 7: Contingency table for the distribution of article presence/absence in *en* subsets according to the definiteness of the antecedent.

	Singular	Plural	TOTAL
<en+ART+que>	106	48	154
<en+que>	72	19	91
TOTAL	178	67	245

TABLE 8: Contingency table for the distribution of article presence/absence in *en* subsets according to the grammatical number of the antecedent.

Since Santana Marrero (2004) asserts that antecedents of time, manner and location are quite common in <en+que> utterances, I examined the influence of prepositional Theta Role on the presence or absence of the article in *en* utterances. Due to the large number of categories in the Theta Role variable, a chi-square analysis cannot be reliably conducted, but I present the descriptive statistics in Table 9. Matching Santana Marrero (2004), we can see that most <en+que> utterances have theta role of Time, Location and Manner (85/91, 93.41%). It is worth highlighting that <en+que> sentences make up more than half of Time (38/68, 55.88%) and Manner utterances (26/33, 78.79%).

	<en+ART+que>	<en+que>	% with article	% without article
Beneficiary	1	0	100	0
Goal	4	1	80	20
Instrument	5	2	71.43	28.57
Location	97	21	82.20	17.80
Manner	7	26	21.21	78.79
State	4	1	80	20
Stimulus	3	0	100	0
Theme	3	2	60	40
Time	30	38	44.12	55.882

TABLE 9: Distribution of article presence or absence per theta role for the *en* subset.

3.4. Discussion

This section summarizes and contextualizes the results (3.4.1) and discusses these results in the context of the broader research project (3.4.2).

3.4.1. Summary of Results

The statistical analysis for this corpus validates the patterns previously identified in the literature and reported in Santana Marrero (2004). Prepositional RCs without an article significantly favour antecedents that are inanimate, singular, and definite, as well as embedded clauses that do not contain negation. The effect of the preposition in-and-of-itself has not been previously examined in the literature, beyond noting that optionality is restricted to a specific set of monosyllabic prepositions.

The preposition *en*, in particular, is worth examining. It is over-represented in the corpus, accounting for over 50% of all utterances, which makes it have a sizeable effect on the results. Post-hoc testing on <*en*+ART+*que*> utterances found a significant interaction between the presence or absence of the article and Definiteness. This interaction followed the pattern of the overall results, with <*en+que*> sentences showing a significant preference for a definite antecedent. However, it is important to note that almost all <PREP+*que*> utterances that have an indefinite antecedent occur when the preposition is *en* (14/15, 93.33%). That is, <*en+que*> utterances seem to be more permissive of the presence of an indefinite antecedent compared with the other prepositions, e.g. (62).

(62) Llegará un momento en que surgirán tensiones...
 arrive.FUT.3SG a moment in QUE emerge.FUT.3PL tensions
 '(There) will come a moment in which tensions will rise...'

Santana Marrero (2004) reports asymmetries between the behaviours of the various prepositions; these differences are also attested in the present corpus. In Table 6 above, we see that the prepositions *en* and *con* have a much higher proportion of <PREP+*que*> utterances (37.15% and 25.49%, respectively) than the other prepositions; they also constitute the majority of <PREP+*que*> utterances, with *en* taking the lion's share. This is in contrast with *a*, *de*, and *por*, which had a small proportion of <PREP+*que*> utterances (7.53%, 0%, and 2.94%, respectively).

3.4.2. What can we do with the preposition effect?

Having identified and quantified the significant effect of preposition in the distribution of <PREP+*que*> utterances, we can begin discussing what this effect can tell us about the structure and properties of prep-RCs. To do so, we must look for commonalities and general patterns between the prepositions that are more friendly to the absence of the article and those that are not. One potential avenue for finding such patterns comes from the relationship between the presence/absence of the article and certain antecedents. Santana Marrero (2004) identifies that <*en+que*> seems to be most common with antecedents of time, place and manner, an assertion supported by the theta role data of this corpus (see Table 9 above). These theta roles are most often assigned to adjuncts, specially in the case of Place and Manner, which are never in argument positions.³⁴ If this distinction extends to the other prepositions, it could point to a deeper distinction underlying this newfound preposition effect.

EN	CON	POR	A	DE
Location (118)	Instrument (23)	Causer (10)	Theme (38)	Theme (18)
Time (68)	Manner (10)	Instrument (10)	Goal (19)	Possessor (7)
Manner (33)			Beneficiary (8)	

TABLE 10: Most common theta roles per preposition

Table 10 shows the most common theta roles associated with each preposition. We can see that the pattern seen for *en* is also found for *con*. That is, the most common theta roles for each prepositional phrase are those associated with the adjunct position, such as Time, Manner and Instrument. In contrast, the most common theta roles for *a* and *de* are associated with DO, IO, and Oblique arguments, such as Theme, Goal and Beneficiary. The exception to this pattern is *por*, which only had one <PREP+*que*> token but has Instrument, associated with an adjunct position, as one of its main theta roles. As has been suggested throughout the

³⁴ This is not to say that a given theta-role is a foolproof indicator of its nominal's syntactic position, but there is an undeniable relationship between the two. For example, AGENT and ACTOR are most commonly found in Subject position, but they might also appear as a prepositional adjunct in passive sentences.

literature, it is likely that *por* favours the presence of the article to prevent ambiguity, i.e., to differentiate the prep-RC from its homonyms *porque* ‘because’ and *por qué* ‘why’.

This distribution pattern suggests that the effect of preposition on the presence or absence of the article might reflect of an underlying difference in the prepositional phrase’s syntactic position. Specifically, it might reflect a contrast between those PPs that enter the derivation as an argument of the embedded clause, and those that enter it as an adjunct. This avenue of exploration is very promising. A’ movement (including *wh*-movement) has been a well-established source of asymmetries between arguments and adjuncts since the earliest days of generative syntax (Ross, 1967), and the study of these differences has triggered innovations in syntactic theory (e.g., Lebeaux, 1991). Since prep-RCs are derived through A’ bar movement in the framework that this project is based on (Cinque, 2020), this avenue of investigation is worth pursuing. Not only does it promise to have very interesting implications should it prove fruitful, but it represents an opportunity to expand this project’s contributions to the syntactic literature beyond what we could contribute by taking the effect of preposition as a contribution in and of itself.

Therefore, I propose the hypothesis that the structural position of the PP in the embedded phrase might condition whether the article is present or absent. That is, I propose that PPs in an argument position prefer the presence of the article, while those in an adjunct position prefer its absence. In the next chapter, this hypothesis is tested through an Acceptability Judgement Task.

4. ACCEPTABILITY JUDGEMENT TASK

While the Corpus has given us very good positive evidence indicating which factors condition article optionality, it is important that we also gather negative evidence of the contexts in which the absence of the article is not permitted. An Acceptability Judgement Task³⁵ (henceforth AJT) is a simple and straightforward way of obtaining this kind of data.

In this chapter, I test the hypothesis that was developed from the results of the corpus study: that the presence/absence of the article is conditioned by the syntactic position³⁶ of the PP within the embedded clause. The first section of this chapter introduces all relevant elements of the study design. Section 4.2 presents the participants and recruitment strategy. Section 4.3 discusses the statistical analysis of the results and Section 4.4 presents some preliminary conclusions.

4.1. Study Design

This section presents the factors relevant to the study design: The conditions and target stimuli design (4.1.1.), predictions of the results (4.1.2), how the stimuli was presented (4.1.3) and the experimental procedure (4.1.4).

4.1.1. Variables

The study has a 2 x 2 x 2 design, with rating on a 7-point Likert scale as the dependent variable and the independent variables being:

- (63) a. Presence/Absence of the article
 b. Definiteness of the antecedent
 c. Syntactic position of the prepositional phrase

The first two variables, Article and Definiteness, are illustrated in Table 11. Each variable has two levels: “Present vs. Absent” for the article (tagged YES and NO, respectively), and

³⁵ This task is also known as grammaticality judgement task, but this name is a bit misleading. Grammaticality is an i-language property that cannot be measured directly. Rather, an AJT asks speakers to self-report their grammaticality judgements, i.e., their instinctive response about the degree to which a given sentence belongs in their grammar (Schutze, 2011).

³⁶ Henceforth, “syntactic position” will be used to refer to the argument/adjunct distinction as an experimental variable.

“Definite vs. Indefinite” for the antecedent definiteness. Each sentence alternated on these two levels, forming minimal pairs.

Article	Definiteness	Stimulus
Present	Indefinite	<i>Los clavadistas admiraron unas plataformas de las que saltaban a menudo</i> The divers admired a.F.PL platforms from ART.F.PL QUE jumped.3PL often
Present	Definite	<i>Los clavadistas admiraron las plataformas de las que saltaban a menudo</i> The divers admired the.F.PL platforms from ART.F.PL QUE jumped.3PL often
Absent	Indefinite	<i>Los clavadistas admiraron unas plataformas de que saltaban a menudo</i> The divers admired a.F.PL platforms from QUE jumped.3PL often
Absent	Definite	<i>Los clavadistas admiraron las plataformas de que saltaban a menudo</i> The divers admired the.F.PL platforms from QUE jumped.3PL often

TABLE 11: Example of stimulus item, ‘The divers admired the platforms from which they jumped often’, modified to form minimal pairs for each of the Article and Definiteness conditions.

To examine the role of syntactic position, stimuli items alternated in terms of the verb. That is, the embedded verb of items in the argument condition is one that selects a Prepositional Phrase as an argument (64a), while verbs in the adjunct condition can have an adjunct PP (64b). The items had otherwise the same linear order.

- (64) a. Los clavadistas admiraron las plataformas **de las que**
the divers admired the.F.PL platforms of ART.F.PL QUE
hablaban a menudo.
talked.3PL often
‘The divers admired the platforms of which they talked often’
- b. Los clavadistas admiraron las plataformas **de las que**
the divers admired the.F.PL platforms of ART.F.PL QUE
saltaban a menudo.
jumped.3PL often
‘The divers admired the platforms from which they jumped often’

To identify whether a given verb selects the PP as an argument or an adjunct, I used some of the diagnostic criteria found in Arce (2017). Arce provides an overview of diagnostic criteria from several sources, such as an elision test and a substitution test. The diagnostic test that

gave me the clearest intuitions is a pseudo-clefting test, originally from Alarcos Llorach (1994, as cited in Arce, 2017). I first constructed simple sentences based on the intended embedded clause, e.g. (65a-b) below for (64a-b) above, respectively. Then, I turned those simple sentences into a pseudo-cleft. If it is an argument, it is possible to create a pseudo-cleft with <PREP+*lo+que*> as the subject (66a). If it is an adjunct, this is not possible (66b), as the adjunct needs a more contentful *wh*-word like *donde* ‘where’ or *como* ‘how’ to head the pseudo-cleft.

- (65) a. Hablaban a menudo de las plataformas.
 talked.3PL often of the platforms
‘They talked about the platforms often.’
- b. Saltaban a menudo de las plataformas.
 jumped.3PL often from the platforms
‘They jumped from the platforms often.’
- (66) a. **De lo que** hablaban a menudo es de las plataformas.
 Of ART.N QUE talked.3PL often is of the platforms
‘That which they talked about often is about the platforms.’
- b. ***De lo que** saltaban a menudo es de las plataformas.
 From ART.N QUE jumped.3PL often is from the platforms

I was able to come up with eight such verb pairs, with four pairs selecting each preposition (see Appendix for full list). Each verb pair was used in two carrier sentences, each with different antecedents, matrix clauses, and utterance-final adverbs. For the verb pair in (64), sentences (67a) and (67b) below represent the second pair of minimal pair sentences. The reason for this is that there were too few viable verb pairs.

- (67) a. El paracaidista compró las avionetas **de las que**
 the skydiver bought the.F.PL small-planes of ART.F.PL QUE
hablaba con gusto
 talked.3SG with joy
‘The skydiver bought the small planes of which he talked joyfully’

- b. El paracaidista compró las avionetas **de las que**
 the skydiver bought the.F.PL small-planes of ART.F.PL QUE
saltaba con gusto
 jumped.3SG with joy
'The skydiver bought the small planes of which he jumped joyfully'

It is worth highlighting that preposition was not directly included as a factor in the stimuli design, despite the results of the corpus analysis clearly showing its influence. The reason for this is that it is not possible to create minimal pairs that alternate in preposition – this would require me to find a verb that selects both PPs as potential arguments, then match it with a verb that selects both PPs as potential adjuncts, in such a way that all four PPs could feasibly have the same referent. I was not able to find one such verb pair, let alone four. Nonetheless, sentences are counterbalanced between the preposition that had the highest proportion of no-article sentences (*en*) and the one with the lowest proportion (*de*), which will allow me to treat preposition as an independent variable in the statistical analysis.

4.1.2. Predictions

Participants are predicted to react to these conditions as follows:

- (i) Sentences where the article is present will be rated significantly higher than those where the article is not present. No significant differences are expected between any subsets of <PREP+ART+*que*> sentences.
- (ii) Following the literature, participants are predicted to rate <PREP+*que*> sentences with a definite antecedent significantly higher than those with an indefinite antecedent.
- (iii) If the syntactic position of the *wh*-phrase conditions the article, then <PREP+*que*> sentences where the prepositional phrase is an argument will be rated significantly higher than those where it is an adjunct.
- (iv) If the effect of preposition does indeed reflect the underlying argument/adjunct distinction (instead of being its own conditioning factor), then <*en+que*> sentences and <*de+que*> sentences will not be rated significantly different from each other.

4.1.3. Lists

The target stimuli consisted of 32 sets with four conditions constructed as in Table 11 (see Appendix 1 for the full list of target items). These sets were sorted into one of four lists using

a Latin Square design. For all items, the antecedent was a F.PL nominal. The main reason for this was to avoid cliticization of *de*. Items in List 1 were semi-randomized and items in the other three lists were sorted manually to the same order.

In addition to the 32 targets, participants saw 32 filler items and 8 removal items. The filler items consisted of object relative clauses. To blend in with the stimuli items, the antecedent of these sentences was always a F.PL nominal and the fillers had one of two conditions: definite antecedent and indefinite antecedent. Each list had an equal number of fillers with each condition. Removal items consisted of *semilibre* RCs (see Section 2.2.1). Half of the removal items were grammatical, and the other half contained errors in number and person subject agreement within the matrix clause. Response files that contained three or more non-target answers in the removal items were dismissed.

4.1.4. Procedure

The study consists of an online acceptability judgement task, created using PsychoPy3 and implemented through its web interface Pavlovia (Pierce et al., 2019). Interested participants received a personalized link to the study on Pavlovia and a consent form, which outlined the task and the conditions for withdrawal. Upon clicking the link, participants were prompted to provide basic demographic information (age, gender and hometown³⁷). Then they saw a page prompting them to click “S” (For *sí*, ‘yes’) if they have read and understood the consent form and are willing to participate. Upon pressing “S”, they saw a series of screens explaining the study mechanics and how to withdraw, before beginning a practice task.

The practice task consisted of three items (one grammatical, one ungrammatical and one grammatical infelicitous), each followed by a feedback screen. Participants were asked to rate each item on a 7-point Likert scale, where “1” is labeled *más inaceptable* ‘more unacceptable’ and “7” is labeled *más aceptable* ‘more acceptable’ (Figure 4). Participants then proceeded to the main task, in which participants saw a series of sentences presented one by one and a Likert scale. Pavlovia collected keypresses and response times. After completing the rating task, participants saw a series of debriefing screens explaining the task and the kinds of data collected, before being prompted one last time to press “S” to submit their

³⁷ Participants were asked for their place of birth (‘hometown’) to exclude participants who are not from Mexico. Nobody listed a place of birth outside of Mexico.

responses. Data of participants who did not complete the task was automatically discarded by Pavlovia. Completed datasets were routinely moved into a secure server.



FIGURE 4: Likert scale presented with the stimulus items.

4.2. Participants

Participants were recruited online through announcements made on a variety of online Facebook groups. The study was open to adult L1 speakers of Mexican Spanish. I targeted university students of the Central Mexico dialectal area, specifically focusing on language and linguistics-related student groups, such as the student clubs of Mexican universities like the *Universidad Nacional Autónoma de México* and the *Universidad del Estado de México*.

The reasons for choosing this dialectal group are twofold: One, since the Spanish of Central Mexico is my L1, using this dialect minimizes the chances of running into confounds that have to do with dialectal differences, such as the use of vocabulary or agreement forms that are unfamiliar to the speaker. Two, Mexican Spanish is not a *dequeísmo* dialect. *Dequeísmo* is a non-standard dialectal feature where the preposition *de* is inserted to select subject and object RCs in contexts that it is not standardly associated with. In dialects with *dequeísmo*, this feature is highly salient and stigmatized, which Santana Marrero (2014) proposes as a potential explanation for the absence of no-article *de* utterances in her corpus. That is, she argues that speakers might overcorrect and avoid using *<de+que>* strings even in contexts where this use would not qualify as *dequeísmo*. Using participants from Mexico eliminates this confound, since they are not expected to have any metalinguistic awareness of *dequeísmo* (Moreno Fernandez, 2016).

Ninety-five potential participants expressed interest and were sent a link to the study on Pavlovia, assigning them a list and a six-digit alphanumeric ID. Of these, 79 successfully completed the study and were sent a \$100 MXN (~\$6 CAD) gift card to Amazon Mexico. Participant files were moved to a secure server, where participant IDs were re-anonymized

and datasets that met the removal criteria were removed. One additional participant was removed due to having a large number of responses with a response time of < 1sec. Finally, each list was trimmed to maintain Latin square balance, so that all lists had an equal amount of datasets. The trimming was done in chronological order, i.e., the last lists to be completed were the ones that were left out.

Having done all the trimming, an equal number of participants from each list were included in the dataset. This resulted in the removal of 7 participants (five from List 1, one from List 2, and one from List 4). The final participant pool (n=64) consisted of 42 women, 21 men and 1 non-binary person and were aged from 18 to 61 years old (mean=26).

4.3. Results

Overall, <PREP+ART+que> sentences had higher mean ratings than <PREP+que> sentences (6.13 vs. 3.86). Looking at the <PREP+que> subsets, those with definite antecedents (4.10) had higher rating than those with indefinite antecedents (3.61), <en+que> (4.74) had higher mean ratings than <de+que> (2.97), and those with adjunct position (3.88) had a slightly higher mean rating than those with argument position (3.83). These mean results were further segmented and analysed in post-hoc testing (see Section 4.3.2).

	Overall Mean	Article Absent	Article Present	Definite Antecedent	Indefinite Antecedent
<PREP+que>	6.13				
<PREP+ART+que>	3.86				
Argument	4.96	3.83	6.08		
Adjunct	5.03	3.88	6.17		
Definite	5.53	4.10	6.22		
Indefinite	4.45	3.61	6.03		
<i>en</i>	5.16	4.74	6.32	5.75	5.31
<i>de</i>	4.82	2.97	5.93	4.58	4.33

TABLE 12: Mean ratings of the target sentences in the AJT, divided by each independent variable and some relevant interactions.

Section 4.3.1 presents the process and results of the statistical modeling using linear mixed effects model comparisons. Section 4.3.2 presents the general post-hoc testing using t-tests and Section 4.3.3 focuses on the post-hoc analysis of potential between-participant differences. Section 4.3.4 justifies the use of a random slope model in the lmer models.

4.3.1. *Lmer*

Results were tested for statistical significance using linear mixed effects model comparisons using the R module *lme4* (Bates et al., 2015). Following the “additive” comparison process described in Winter (2013), I first compared a null hypothesis model with Participant and Item as random effects (**Rating~1+(1|PartID)+(1|ItemID)**) to models containing one fixed effect factor. In addition to the three main factors built into the study design (Article, Definiteness and Syntactic position), the effect of Preposition was also investigated. The models that had significantly more explanatory power than the null model were those containing a single fixed effect of Article ($\chi^2(1)= 124.51, p<0.001$) or Preposition ($\chi^2(1)= 19.289, p<0.001$). Models containing Syntactic Position ($\chi^2(1)=0.072, p=0.789$) or Definiteness ($\chi^2(1)=1.854, p=0.173$) as their sole fixed effect did not have significantly more explanatory power than the null model.

I then compared a one-factor model with Article as the fixed factor to models with two fixed factors. The two-factor models that included Preposition ($\chi^2(1)=59.71, p<0.001$) or Definiteness ($\chi^2(1)=4.992, p=0.025$) as an independent fixed effect had significantly more explanatory power than the one-factor model. The two-factor model containing Syntactic Position ($\chi^2(1)=0.192, p=0.662$) did not. The two-factor model that contained an interaction between Article and Preposition had significantly more explanatory power than the model where the factors were independent ($\chi^2(1)=35.145, p<0.001$).

This two-factor model with an Article:Preposition interaction was compared to models with a third fixed factor. The three-factor model that included Definiteness as an independent fixed effect had significantly more explanatory power than the two-factor model ($\chi^2(1)=10.817, p=0.001$); the model that included Syntactic Position did not ($\chi^2(1)=0.406, p=0.524$). This three-factor model with an independent Definiteness effect also had more explanatory power than a model that included Definiteness interactions ($\chi^2(3)=4.525, p=0.210$) and a four-factor model ($\chi^2(1)=0.442, p=0.506$).

Having determined the random intercept of the model, I created the most extensive random slope that was justified by my design (following Winter, 2013).³⁸ The random slope

³⁸ I chose to add a random slope only to the Subject factor, but not the Item factor. I did not add a random slope to Items because it was not clear to me that doing so was justified by my experimental design, as I had not

model (68) had significantly more predictive power than the random intercept model ($\chi^2(14)=271.33, p<0.001$), and was taken to be the best-fit model.

(68) Rating ~ Article * Preposition + Definiteness + (1 + Article * Preposition + Definiteness | ParticipantID) + (1 | ItemID)

The lmer algorithm defines the intercept by creating a “default group,” alphabetically. In this case, the default is a sentence with a definite antecedent, no article and the preposition *de*. The best-fitting model predicts that sentences in these groups would get an approximate rating of 3.148 ± 0.194 . All other factors being held constant, a sentence with the article is predicted to get a higher rating by 2.955 ± 0.212 . All other factors being held constant, a sentence with the preposition *en* is predicted to get a higher rating by 1.762 ± 0.188 . All other factors being held constant, a sentence with an indefinite antecedent is predicted to get a lower rating by -0.346 ± 0.112 . There is predicted to be a wider difference between sentences with and without the article for sentences with the preposition *de*, compared to the difference in ratings for sentences with the preposition *en* (-1.367 ± 0.228). The lmer report is shown in Table 13.

	Estimate	Std. Error	t-value
(Intercept)	3.1475	0.1935	16.269
ArticleYES	2.9551	0.2119	13.942
PrepEN	1.7617	0.1879	9.377
DefinitenessIND	-0.3457	0.1121	-3.085
ArticleYES:PrepEN	-1.3672	0.2283	-5.988

TABLE 13: Fixed effects for the best-fit model.

4.3.2. Initial post-hoc testing

I ran t-tests on all subsets of the data that were deemed significant by the lmer modeling and applied a Boniferroni correction (see Figure 5 for the means reported by these tests). For all subsets, sentences with the article were rated significantly higher than those without the article, as predicted in the model. Sentences with *en* were rated significantly higher than

predicted any variability between items that was not tied to the fixed factors, and there is no overlap between conditions. On the other hand, I did expect individual participants to vary in their rating of different conditions.

those with *de*. While this matches the predictions made by the model, it is worth highlighting that this difference was maintained even when the article was present, with a mean of 6.32 for $\langle en+ART+que \rangle$ sentences and 5.93 for $\langle de+ART+que \rangle$ sentences ($t(909.8) = 4.45$, $p < 0.001$). Subsets in the definite antecedent condition were rated higher than those with an indefinite antecedent, but this difference was not considered significant in most subsets. The notable exception to this trend is within the no-article condition. For $\langle en+que \rangle$ sentences, those with definite antecedents (5.09) were rated significantly higher than those with indefinite antecedents (4.38) ($t(508.39) = 4.16$, $p < 0.001$).

MEAN RATINGS

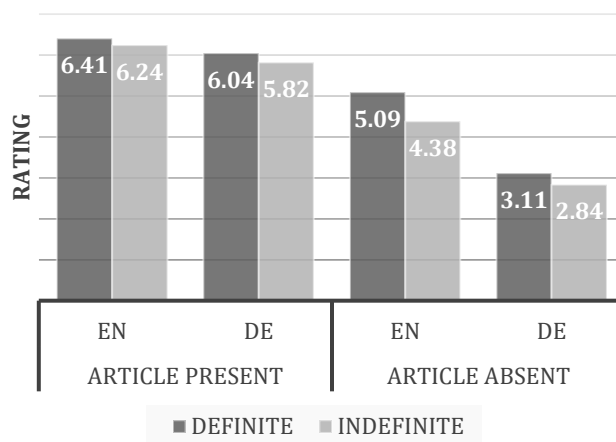


FIGURE 5: Mean ratings of each significant subset of conditions.

I conducted visual inspections on the keypresses for all items, by subset and by individual means, to detect potential outliers that might be skewing the results. There were a few specific items that were rated quite low relative to other items in that condition, but these items did not follow a specific pattern that might affect the average rating for any given factor. The main reason for these lower ratings seems to be attachment ambiguity, i.e., items where the adjunct could modify either the matrix or the embedded verb. For example, in (69), the adverb *felizmente* 'happily' is closest to the embedded verb *creían* 'believed (in)'. However, semantically it would make more sense for it to modify *defendieron* 'defended'. Items with non-human subjects in the matrix clause were also rated lower sometimes, presumably because of their Agent theta role, but this was not consistent across conditions nor minimal pairs.

- (69) Los gerentes defendieron las compañías en las que
 the managers defended. 3PL the.F.PL companies in ART.F.PL QUE
 creían felizmente
 believed.3PL happily
'The managers defended the companies in which they believed happily'

The subset keypresses also showed the overall patterns identified by the lmer model, but there seems to be a lot of variation within the no-article conditions. For example, the left-hand chart in Figure 6 shows that the most common rating given to *de* items is one (157 keypresses), but ratings are distributed across the scale, and there are instances where *<de+que>* items were given the highest rating (27 keypresses). The bar plot for the no-article *en* items (Figure 6, right-hand side) can be seen as a mirror of its counterpart,³⁹ with seven (127 keypresses) being the most common rating and one (50 keypresses) being among the lowest. However, we see a lot of variability within the *en* ratings. The second most common rating is five (100 keypresses) rather than six, and the keypresses for ratings one through four are comparable to each other. By contrast, the *de* plot forms a smooth descending curve. The variability of the *<en+que>* distribution, as well as the fact that the lmer model contains a random slope, suggests that there might be strong differences between participants.

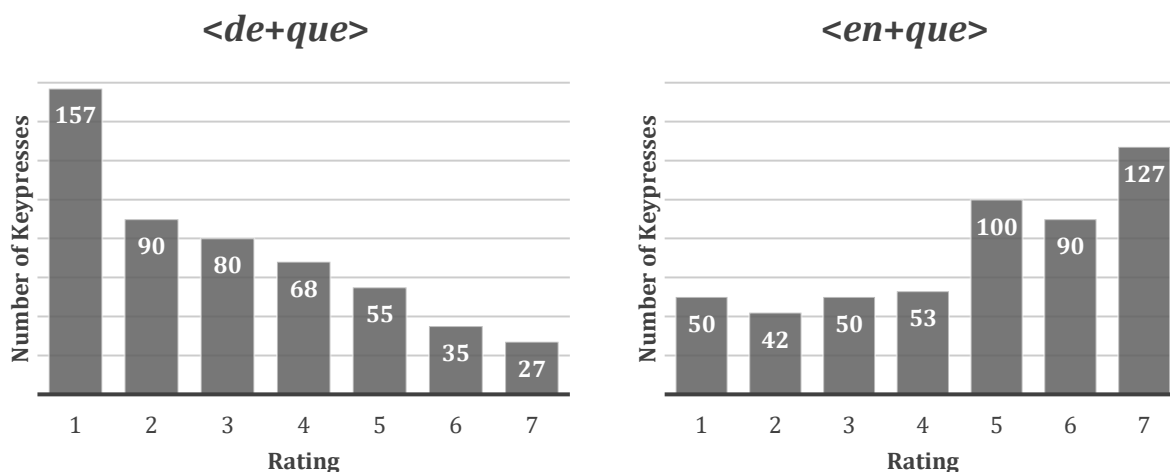


FIGURE 6: Bar chart of total keypresses per preposition for the no-article conditions.

³⁹ This symmetry is maintained if we take acceptable vs. unacceptable judgements in aggregate (keypresses 5-7 and 1-3, respectively). *<de+que>* sentences received 117 acceptable and 327 unacceptable ratings; *<en+que>* received 317 acceptable and 142 unacceptable ratings.

4.3.3. Analysis of between-participant differences

To identify whether the differences in participant responses could be sorted into distinct population groups, I first conducted a visual inspection of per-item participant intercepts using the *coef()* function of the *lme4* R package. The model showed a wide range of participant intercepts, from just over 1 to 5. This suggests that, while most people gave drastically different ratings to *<de+que>* and *<en+que>*, some participants did not, rating both sets as either grammatical or ungrammatical.

To begin teasing apart these potential populations, I ran t-tests on each participant's ratings, comparing their no-article ratings to *en* vs. *de*. This was done through an R script. Of the 64 participants, 34 rated no-article sentences with *en* significantly higher than those with *de* (**p<0.05**). I sorted the remaining participants into two groups: the "optional article" group, who rated both types of no-article sentences as relatively acceptable (*de* rating ≥ 4 and *en* rating ≥ 4), and the "mandatory article group", who rated both types as relatively unacceptable (*de* rating < 4 and *en* rating < 4). There were 15 participants in the "optional article" group and nine in the "mandatory article" group.

In addition, there were six participants who rated *en* sentences as more grammatical than *de* sentences, but where the difference between ratings did not constitute a significant difference. For some of these participants, this was because they rated both types of no-article sentences quite close to the median. For example, participant *89v71N6EMUB4* had an *en* mean of 4.625 and a *de* mean of 3.625 (**t(12.12) = -1.106, p=0.290**). However, this was not always the case, such as for participant *86wz52eA3VJ2*, who had an *en* mean of 5.5 and a *de* mean of 2.875 (**t(13.99) = -2.061, p=0.058**).

The results of the per-participant t-tests suggest that, while there is variation with regards to how participants react to the *<PREP+que>* condition, this variation might not be as easily captured as I had expected it to be. It is worth highlighting the fact that the per-participant t-tests were dealing with a very small amount of data ($n=8$), and are likely less reliable than a larger sample would be. Nonetheless, the results do show enough of a trend that I feel comfortable suggesting that future research on these three speaker types could be fruitful.

4.3.4. Lmer modeling with random slopes

Whereas a standard lmer model (aka. a random intercept model) assumes that the effect of the fixed factor applies equally to all participants and/or items, a random slope model adjusts fixed factor and interaction slopes for every individual member of the random factor for which it is specified. To illustrate this, I have selected the coefficients of a participant from each of the three populations I have tentatively identified above: the “optional article” group (i), who treat all <PREP+*que*> sentences as acceptable, the “mandatory article” group (ii), who rate all <PREP+*que*> sentences as unacceptable, and the “distinguisher” group (iii), that rate <en+*que*> sentences as acceptable and <de+*que*> sentences as unacceptable. I focus on the YES ratings (‘article present’). Coefficients have been rounded down to three decimal points.

Group	PartID	Intercept	YES	EN	IND	YES:EN
(i)	56qoh2l0Y74k	4.562	1.466	1.477	-0.184	-1.181
(ii)	5EJ7DH8K17C4	1.755	4.992	1.096	-0.257	-1.028
(iii)	5K631Jrrt3u0	3.115	2.541	2.100	-0.344	-1.465
	Fixed coefficients	3.1475	2.955	1.762	-0.346	-1.367

TABLE 14: Sample of the coefficient report for the random slope lmer algorithm.

As stated above, the “default group” was defined as a sentence with a definite antecedent, no article and the preposition *de*. In a random intercept model, the program calculates individual default intercepts for each participant (and item) and adjusts the slope by the fixed coefficient amount. In a random slope model, these adjustments are done individually, which is crucial when dealing with a range of behaviour.

For example, the model predicts that participant (i) rates default sentences at 4.562 on the 7-point Likert scale. Applying default adjustments, the random intercept-only model predicts that they rate ‘YES sentences’ (i.e., <de+ART+*que*>) 2.955 points higher than the default rating. In this case, this leads to an impossible prediction of a rating of 7.517 on a 7-point Likert scale. By contrast, the random slope model adjusts the intercept by 1.466, predicting ‘YES sentences’ to be rated at a much more realistic 6.028.

For participant (ii), who is predicted to rate the default sentences at 1.755, the intercept model would also adjust the ‘YES sentences’ by 2.955 points, predicting them to be rated at

4.710. While not impossible, this would be a surprisingly low rating for what were intended to be the unmarked stimulus items. In the random slope model, this is adjusted by 4.992, leading to a more expected prediction of 6.747 for the ‘YES sentences’.

Participant (iii), who has a fairly average intercept, has similar predictions in both models (For the ‘YES sentences’, 6.070 in the random intercept model and 5.656 in the random slope model).

Despite the variation in individual adjustments, the general trend matches that of the fixed effect estimates. That is, all other things held constant, sentences with the article are rated higher than those without the article, sentences with *en* are rated higher than those with *de*, and sentences with an indefinite antecedent are rated lower than those with a definite antecedent.

4.4. Preliminary Discussion

The analysis of the acceptability judgement task indicates that sentences with an article are rated significantly higher than those without an article; sentences with the preposition *en* are rated significantly higher than those with the preposition *de*; and sentences with a definite antecedent are rated higher than those with an indefinite antecedent. We also saw a significant interaction between article presence and preposition, indicating that *<en+que>* sentences are rated significantly higher than *<de+que>* sentences.

The prediction that we would not see any significant differences within any pair of subsets that contained an article did not pan out, as *<en+ART+que>* sentences were rated significantly higher than *<de+ART+que>*. However, we must highlight that in both cases, the average rating for these items was around ‘6’, and that these items are still firmly considered as grammatical.

The prediction that *<PREP+que>* sentences with a definite antecedent would be rated significantly higher than those with an indefinite antecedent is also supported by the results of the study, although in a less uniform manner than expected. Definiteness had a significant impact on the rating of *en* sentences, but not on *de* sentences. This result reinforces the interaction between definiteness and the preposition *en* that we identified in the corpus analysis.

The main prediction underlying the design of this task was the role of syntactic position in the acceptability of <PREP+*que*> sentences. This prediction was not supported by the data. Syntactic position was not a significant predictor of Rating, and any model that included position as a factor gave significantly worse predictions than those without it. Rather, it appears that the Preposition itself influences whether the absence of the article is considered grammatical. I had expected this effect to disappear as the effect of syntactic position became highlighted, but that is not what happened. Despite between-participant variation, it is clear that sentences without an article are considered more acceptable when the preposition is *en*.

The results of this task indicate that preposition and definiteness have a significant effect on the acceptability of <PREP+*que*> sentences. Having learned all I can from this task, I can now return to the theoretical analysis of article optionality. In the next chapter, I discuss how the results of this task can be integrated into an updated understanding of the principles behind article optionality.

5. DISCUSSION

This chapter brings together all of the components of this research project, contextualizing the results of the experimental tasks into our theoretical understanding of prep-RC structure. My primary goal is to update the current analysis of how prep-RCs with and without the article are derived, integrating our current knowledge of these structures into Cinque's (2020) framework. I examine some of the syntactic properties that underlie the presence or absence of the article in Spanish prep-RCs, focusing mainly on the effect of definiteness.

Section 5.1 provides an overview of the results of the experimental tasks, offering some initial thoughts on the implications of these findings and justifying the choice to focus on definiteness and preposition as predicting factors. Section 5.2 discusses how the works of Brucart (1992, 1994, 1999) and Cinque (2020) contribute to my analysis of the structure of prep-RCs with and without the article, specifically as it pertains to antecedent definiteness. Section 5.3 presents the derivations for both structures. Section 5.4 elaborates on how my proposal diverges from Cinque's (2020) universal double-headed structure. Section 5.5 explores some potential explanations for the effect of preposition on the presence/absence of the article.

5.1. Overview of results

In this section, I discuss the results of the corpus study (Chapter 3) and the acceptability judgement task (Chapter 4) in the context of what these results might be able to tell us about the syntactic properties of the article and the structure of prep-RCs.

The corpus study identified that the following factors had a significant correlation with the absence of the article: an antecedent that is definite, inanimate, and singular, lack of negation in the embedded clause, and specific prepositions. I also found an interaction between definiteness and the preposition *en*, where indefinite antecedents, while still being a minority, are somewhat permissible with *<en+que>*. That is, almost all instances of indefinite antecedents in a *<PREP+que>* utterance were *en* sentences. These results formed the basis for the stimulus design of the AJT, where I tested the acceptability of presence vs. absence of the article, definiteness of the antecedent, and syntactic position (argument vs. adjunct). I also looked at two prepositions that I thought representative of the propensity

towards accepting vs. rejecting the absence of the article, *en* vs. *de*. Syntactic position was not a significant predictor of rating, but participants had a very marked preference towards the presence of the article, and towards the preposition *en*. The preference for definite antecedents was less dramatic, but statistically significant.

By itself, the large difference in rating between sentences is quite valuable in that it supports the overall premise of the linguistic literature, that is, that the article is not quite optional. It also specifically challenges Suñer's (2001) claim that the article is a prescriptive artifact. Had that been the case, <PREP+*que*> sentences would have been categorically rejected, but most participants actually distinguished between different conditioning factors. Although we did see some participants who categorically rejected or accepted the absence of the article, a majority of the participants distinguished between *en* and *de*, giving *en* sentences significantly higher ratings. Overall, <*en+que*> sentences were rated as marginally acceptable, whereas <*de+que*> sentences were rated as unacceptable.

Post-hoc testing indicated that the difference between definite and indefinite antecedents was only significant within <*en+que*> sentences. This echoes the results of the corpus study, where almost all of the <PREP+*que*> sentences that had an indefinite antecedent were <*en+que*> sentences. I suggest that the reason for this is that, since participants were categorically rejecting <*de+que*> sentences, any nuance that was brought in by the definite antecedent was overlooked. By contrast, <*en+que*> sentences had a higher acceptability. This allowed the participants to make more nuanced distinctions, allowing the definiteness effect to come through. In the following sections, I focus on the behaviour of <*en+(ART)+que*> sentences in order to better observe the nuanced effect of definiteness.

5.2. Grounding the definiteness data in the theory

In this section, I examine the results of the experimental tasks in the context of the theoretical works introduced in Chapter 2. I investigate whether these frameworks can account for the factors I present above. In Section 5.2.1, I discuss how Brucart (1992, 1994) provides a potential explanation for the definiteness effect, where <*en+que*> sentences were rated significantly higher when the antecedent was definite. In Section 5.2.2, I explore in more general terms whether this distinction can be captured through Cinque's (2020) theoretical framework.

5.2.1. *Brucart and the definiteness effect*

Brucart (1992, 1994) identifies both definiteness and preposition as some of the factors that influence the presence or absence of the article. I focus on definiteness in this section, and return to what Brucart (1992, 1999) has to say on the effects of different prepositions in Section 5.5.

After making the simple observation that the article is only absent when there is a definite antecedent, Brucart (1992, 1994) suggests that definiteness could be the underlying principle behind the polarity constraint (the constraint that a <PREP+*que*> sentence can only occur when there is no negation in the embedded clause). Clauses that contain negation are weak islands, meaning that they are subject to certain restrictions in terms of which operator phrases can undergo A-bar movement from within them (Rizzi, 1990, as cited in Brucart, 1994). Specifically, Cinque (1990, as cited in Brucart, 1994) claims that operator phrases that are “non-referential” are subject to this island constraint. Since the article seems to be mandatory in negated prep-RCs, Brucart (1994) proposes that the article must be providing that referential property to the operator, allowing it to undergo A-bar movement.

Here I must take a moment to clarify what Cinque (1990, as cited in Brucart, 1994), and therefore Brucart (1994), means by “non-referential”, since referentiality is not a syntactic property in-and-of-itself. Cinque (2020) uses the terms “non-referential” and “non-specific” interchangeably, and Brucart (1994) discusses “referentiality” in the context of definite vs. indefinite antecedents. Therefore, I will proceed with the assumption that Brucart was talking about specificity.

While the details and implications of polarity as a restricting factor are beyond the scope of this research project, the crucial idea is that the article is providing a [+specific] feature to the operator phrase, allowing it to overcome the island effect in negated sentences. I take this idea that the article is [+specific] and incorporate it into my proposal in Section 5.3.

5.2.2. *Updating my analysis through Cinque (2020)*

As I mention in Section 2.3.2, it is remarkably straight forward to map out the <PREP+*que*> structure into Cinque’s (2020) two-headed structure, specifically the derivation where the

external head (i.e., the antecedent) is the overt one. In this section, I explore whether we can make sense of the <PREP+ART+*que*> construction under this same framework.

If we assume Cinque's structure, where the internal head is a standard nominal phrase, we could frame the optionality issue as a question of what motivates deletion of the entire nominal in some cases (<PREP+*que*>), but partial deletion/deletion of the bare nominal in others (<PREP+ART+*que*>). Following the core assumption that the internal and external heads are categorially identical, and that this identity (what Cinque (2020) calls "non-distinctness") is a prerequisite for deletion, the only possible answer is that the external head is a bare nominal. This would mean that the site of attachment for the relative clause is smaller than dP. The claim that a <PREP+*que*> clause and a <PREP+ART+*que*> clause have different sites of attachment is not one I am willing to make. It represents a huge departure from the standard analysis with no obvious benefits. Most saliently, it would imply that a clause containing a <PREP+*que*> and one containing <PREP+ART+*que*> are different semantic types. To the best of my knowledge, no meaning differences have ever been attributed to the presence or absence of the article.⁴⁰ In the absence of such differences, this possibility can be discarded.

Another possibility to fit <PREP+ART+*que*> under Cinque's model is to look at how Cinque (2020) analyses prep-RCs in Italian. Italian, like Spanish, is a language with mandatory pied-piping. It is characterized by the *wh*-word *cui*, which Cinque identifies as a relative operator equivalent to the English *which*. Let us remember that Cinque defines non-distinctness on a categorial level. That is, the internal and external heads do not need to be identical in form, they only need to co-refer and be the same type of category phrase. For restrictive RCs, this category is the intermediate nominal phrase that we have called dP, the one immediately below the site of attachment of the RC. Cinque argues that, since the external head in Italian prep-RCs does not trigger the deletion of the internal head *cui*, that indicates that *cui* must be categorially distinct from the antecedent.

⁴⁰ In fact, some authors have explicitly stated that there is no such difference, emphasizing that article alternation is "a purely syntactic phenomenon" (Blas & Vellón, 2018, p. 6) and that the choice to use the article in contexts that permit <PREP+*que*> is "purely aesthetic" (Porto Dapena, 1997, p. 24).

Cinque generalizes that, in cases where the internal head is categorially larger than the external head (i.e., a DP or KP), the internal head is not deleted. Instead, it can be realized “by a *wh*-pronoun or *wh*-phrase, or by a resumptive pronoun or epithet.” (Cinque, 2020, p. 37). If we assume that the article in <PREP+ART+*que*> sentences is one of these categorially larger internal heads, that would allow us to capture the distinction between <PREP+*que*> and <PREP+ART+*que*>. ⁴¹ In the next two sections, I walk us through the derivations of these two sentence structures and outline the ways in which my proposal differs from both Bruccart’s and Cinque’s work.

5.3. A structural account of article alternation

As mentioned in the previous section, Cinque (2020) proposes that pronominal operators that do not undergo deletion have larger nominal structures than those that do. This fits well with Bruccart’s (1992, 1994) notion that the article provides a more definite quality to the operator phrase, which I have interpreted as the [+specific] feature. If we assume that syntactic specificity is located at the D-head, then we can say that the operator phrase in the <PREP+ART+*que*> structure is (at minimum) a DP, as it is valued for specificity. I now present the derivations for each structure using (70), one of the stimulus items, as my example.

- (70) La joven quemó las cartas en (las) que pensaba
 the youth.F.SG burned the.F.PL letters about ART.F.PL QUE thought.3SG
 demasiado
 too-much

‘The young woman burned the letters that she thought about too much.’

⁴¹ To maintain continuity in my terminology, I will continue to refer to the article as such, but this should not be taken as a categorization claim. Other than a definite article, it is also possible to characterize this particle as a pronominal clitic, or part of the pronominal operator phrase.

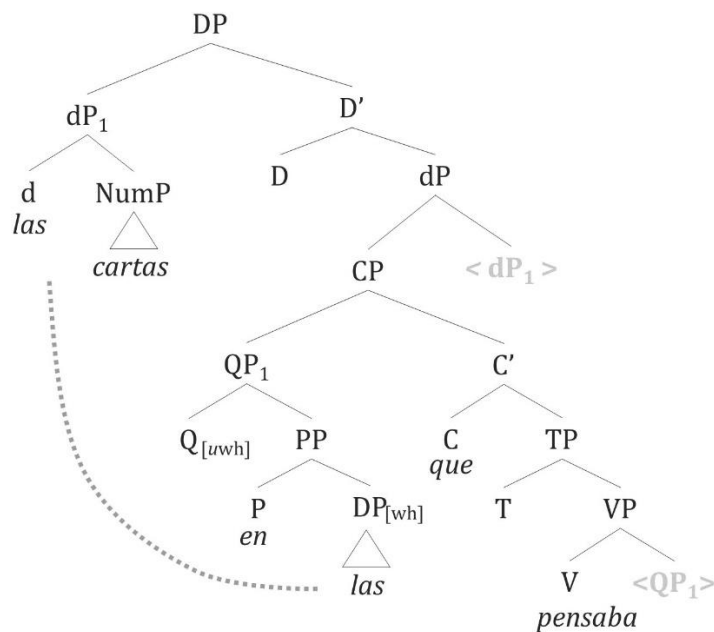


FIGURE 7: Tree representing the <PREP+ART+*que*> structure.

Figure 7 represents my proposal for the <PREP+ART+*que*> structure. The operator phrase [_{DP} *las*],⁴² as well as the PP that contains it, are selected by a null Q head. The QP then moves to Spec, CP. The CP modifier adjoins to an existing dP, '*las cartas*', which subsequently moves to Spec, DP.⁴³ As we can see, the dP is in a c-commanding position over the operator phrase, but since it is not of the same category, it does not trigger deletion. This derivation should be available to the speaker as long as the external head is smaller than the article, regardless of its form or features, which explains why this derivation has the widest distribution.

⁴² The operator phrase is represented as a triangle to avoid making any claims on the categorization of the article and the components of this nominal phrase beyond the category of its highest level. Brucart (1992, 1994) represents this structure as an article selecting a null operator ([*las* + OP]), but it could also be possible for the [*wh*] feature to be hosted elsewhere in the nominal tree.

⁴³ Cinque (2020) claims that the dP's upward movement is independently motivated but does not explain the underlying theory that backs this assumption. Validating this claim is beyond the scope of this thesis, but it is worth mentioning that, as part of his proposed account of nominal modification in Spanish, Fabregas (2017) claims that all modifiers with a restrictive interpretation are post-nominal in Spanish. In tune with Cinque's claim, he proposes that this post-nominal position is derived by a smaller nominal head moving up beyond the merging point to derive that post-nominal word order.

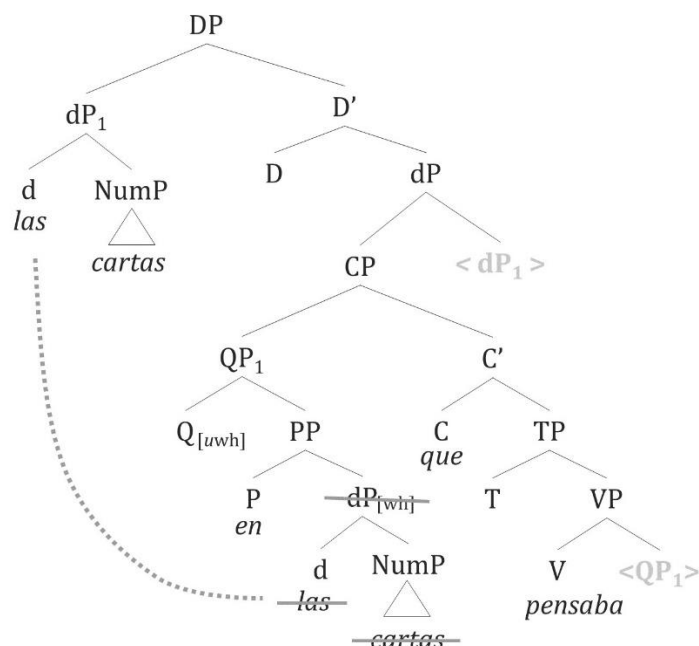


FIGURE 8: Tree representing the <PREP+*que*> structure with definite antecedent.

By contrast, in the <PREP+*que*> structure (Figure 8), the preposition *en* selects the dP ‘*las cartas*’. I assume that this dP contains the same operator feature as [_{DP} *las*], allowing the PP to be selected by Q^0 and moved to Spec, CP. The CP modifier adjoins to an existing dP ‘*las cartas*’, which subsequently moves to Spec, DP. Same as in Figure 7, this puts the external dP in a c-commanding position over the operator phrase, which triggers deletion under identity.

In theory, this mechanism is not dependent on the definiteness of the antecedent. To put this in concrete terms, the trees above should look the same regardless of whether the antecedent is *las cartas* or *unas cartas*. However, in practice, many speakers seem to have a strong preference to allow the <PREP+*que*> structure only when the antecedent is definite. Here, we can turn back to Bruccart (1992) for a possible explanation. Even if it does not contribute to the derivation of the prep-RC, the article provides additional evidence of the operator’s existence and position in the structure, making the co-reference chain overt rather than covert. This redundancy allows the prep-RC to be interpreted more easily.

5.4. Outstanding issues

In this section, I address the ways in which my proposal departs from Cinque’s (2020). The first, most immediately evident difference is that I have incorporated Cable’s (2010) Q-

Theory as part of the relative clause structure. The purpose of this addition is to fill a gap in Cinque (2020). The internal head of Cinque (2020), which outwardly appears to be a regular nominal phrase, replaces what we would call a null operator in standard theory. However, Cinque (2020) does not make it clear what motivates A-bar movement to Spec, CP in the absence of an operator. The movement of a pied-piped phrase is even harder to motivate under this framework. The introduction of Q-Theory and the QP level addresses those issues in a simple and straightforward way, without contradicting or interfering with the mechanisms behind Cinque (2020).

Another significant difference is the way that I have interpreted the syntactic structure of the internal and external heads. Cinque's double-headed structure contains two phrasal layers between dP and DP, FP and YP. As their sole function seems to be to serve as attachment site for the RC (in the case of YP) and a landing site for the dP (in the case of FP), I am not including them in this proposal, instead attaching the RC as a dP adjunct and assuming that dP moves to Spec, DP.

I also must address the concept of "dP". Cinque uses this term as a placeholder name for "the portion of the nominal extended projection including weak determiners below the Merge position of the RC." (Cinque, 2020, p. 41) Instead, I have chosen to treat dP as an actual layer of nominal structure between NumP and DP.

Cinque (2020) situates the article in DP, above the site of attachment. However, I assume that the definite article is located below the site of attachment, as part of the dP. Specifically, I assume that dP contains the [\pm definite] feature that triggers the presence of an overt article; DP contains the [\pm specific] feature instead. My reasoning for this is two-fold. First, this assumption prevents overgeneration. If we say that dP is a bare nominal, we would expect bare nominals to freely enable the absence of the article. However, <PREP+*que*> is not any more likely to occur with a bare nominal than it is with other types of indefinite nominal. That is, while exceptions exist, e.g. (71), there is a consistent trend of rejecting and not producing <PREP+*que*> with indefinite antecedents, including bare antecedents.

- (71) ...hay días en que me gano menos de siete mil...
 BE.EXT days in QUE CL.1SG earn.1SG less than seven thousand
'There are days in which I earn less than seven thousand...'

(Corpus del Español)

Furthermore, there is independent evidence from French that shows that definite articles in Romance languages are not Determiners. Déchaine & Wiltschko (2002) proposes that definite articles in French have a smaller nominal structure than those in English, as they can have both a specific (72i) and a generic (72ii) interpretation. This indicates these features must not be inherent to the article, but rather hosted higher up in the nominal structure. We can see in (73) that some Spanish nominals also have this double interpretation.

(72) Jean aime le vin.

Jean likes DET wine

= i. *'Jean likes the wine.'*

= ii. *'Jean likes wine.'*

(Déchaine & Wiltschko, 2002, p. 430)

(73) A Juan le gusta el vino

A Juan CL.3SG enjoy ART.M.SG wine

'Juan likes the wine' / 'Juan likes wine'

This claim, and its applicability to Spanish, opens the possibility that the internal head in <PREP+*que*> structures can contain the article without necessarily being a DP. This intermediate structure, higher than the level of structure that would trigger a generic interpretation but lower than DP, is what I propose as dP.

5.5. Can we ground the preposition effect in the theory?

Both the acceptability judgement task and the corpus study showed that the choice of preposition has a significant effect on the presence or absence of the article beyond what had been theorized in the literature. It has long been known that only a handful of prepositions⁴⁴ can appear in the <PREP+*que*> structure, but even within this group I observed consistent and significant differences. The AJT found that <*en+que*> was rated significantly higher than <*de+que*>; a post-hoc visual inspection showed that even for participants for whom this distinction was not significant, <*en+que*> was still rated higher. In this section, I take a look at some possibilities that could explain the differences between prepositions *en* and *de*.

⁴⁴ These are *a*, *con*, *de* and *en*; most sources also identify *por*, but see Martinez (1989) for an exception.

One of the factors that Brucart (1994, 1999) identifies as mediating the presence/absence of the article is what he calls “lexical properties”. Let us recall that Brucart (1994) proposes that Spanish prep-RCs have a requirement to transmit φ -features between the antecedent and the operator. The role of the article is to serve as an intermediary between the antecedent and the operator. Since this intermediary is only needed in prep-RCs (as opposed to other types of RCs), this must mean that the preposition interferes with the relative operator reading the φ -features of the antecedent.

Brucart speculates that different prepositions must have different degrees of “opacity”, i.e., each preposition might differ in terms of how much they interfere with the feature transmission. Since polysyllabic prepositions seem to always require the article, they must have the highest opacity. If we follow his argument, we could speculate that *en* is less opaque than *de*, and then it is our job to figure out a) what this property of “opacity” actually is and b) if there are any evidence for or proposals about *en* and *de* having different properties that might result in this difference.

For question (a), I do not have an answer. For question (b), Ursini (2013) proposes that monosyllabic prepositions may have a variety of underlying syntactic structures that relate to both their meanings and the ways in which polymorphemic prepositions are formed. While the focus of this analysis is on spatial prepositions, what is interesting for our purpose is that he depicts *en* as the specifier of a null preposition head. By contrast, the structures proposed for the various *de* heads merge as prepositional heads. This null preposition could be posited to have the lowest “opacity”, enabling feature transmission to a higher degree. However, without a better understanding of the structural properties of prepositions and their role in the operator phrase, this idea can only remain as speculation.

Something else that Brucart (1999) mentions as a “lexical property” is that there seem to be a small number of antecedents that have a strong preference for $\langle en+que \rangle$, such as *modo*, *manera* and *forma* ‘way/manner’, e.g. (74).⁴⁵ The reasons why this is the case are beyond the scope of this research project (Brucart speculates they might actually not be

⁴⁵ From a visual inspection, the corpus shows that most of the sentences with these antecedents were part of $\langle \text{PREP}+que \rangle$ utterances. It is also worth mentioning that these kinds of antecedents were not included in the AJT stimuli.

restrictive RCs, but rather something predicational), but it is worth mentioning that these antecedents were quite common in the corpus study. It is possible that Spanish speakers might be extending the pattern of these common phrases, accepting *<en+que>* through analogy even in contexts where it is not quite grammatical and/or harder to process.

- (74) ...información pertinente a **la forma en que** se debe llenar
 information relevant to the.F.SG way in QUE PSV must fill.INF
 el contrato laboral
 the contract labour

'...relevant information on the way in which the employment contract must be filled'

(Corpus del Español)

Conversely, it could also be the case that the ubiquity of contexts in which *<de+que>* would be ungrammatical might lead speakers to reject *<de+que>* sentences even in contexts where it should be grammatical. As Ursini (2013) mentions, *de* is quite prevalent in polymorphemic prepositions such as *delante de* 'in front of', *a través de* 'across from', *después de* 'after'. Since *<PREP+que>* is not possible with these polymorphemic prepositions, speakers might come to reject *<de+que>* across the board.⁴⁶

In this chapter, I outlined some promising possibilities for what assumptions we might reasonably derive from the data collected. Having followed the argumentation for these possibilities as far as I can take them, in the next section I outline the overall contributions of this thesis and propose possible avenues for future work.

⁴⁶ An intriguing aspect of this issue, which is unfortunately beyond the scope of this paper, is the historical trajectory of the presence/absence of the article, in which the *<PREP+ART+que>* has been gaining prominence over the original *<PREP+que>* structure. An analysis of 18th Century texts (Guzman Riverón, 2012) shows that *de* has historically patterned with *en* and *con* in that the *<PREP+que>* form was the most prevalent, but both Santana Marrero (2004) and my own corpus found very few instances of *<de+que>*. Even the descriptive sources in Section 2.2.2., which were written 20 to 30 years ago, present *<de+que>* example sentences that I found awkward.

6. CONCLUSION

The primary goal of this thesis project was to examine the phenomenon of article optionality in Spanish prepositional relative clauses in as much breadth and depth as possible. It centered around the following research questions: (i) What are the syntactic and semantic factors that influence the presence or absence of the article? (ii) How does the article fit within the syntactic structure of the Spanish prep-RC?

To answer the first question, I conducted a corpus study and an acceptability judgement task that provided positive and negative evidence, respectively, of these factors. The corpus study found that the following factors were significant predictors of the presence of the article: an antecedent that is animate, indefinite, or plural; negative polarity in a sentence; which preposition selects the relative operator phrase. Post-hoc testing also found a significant interaction between Definiteness and the preposition *en*. I hypothesized that the Preposition factor might actually reflect an underlying structural distinction (whether the PP was an argument or an adjunct of the embedded clause), and designed an acceptability judgement task to test that hypothesis. In this task, <PREP+ART+*que*> sentences were rated significantly higher than <PREP+*que*> and <*en*+(ART)+*que*> sentences were rated significantly higher than <*de*+(ART)+*que*>; the argument/adjunct distinction was not significant. There was a small but significant effect of Definiteness; post-hoc testing found that this distinction was only significant for <*en*+*que*>.

These results provide empirical evidence in support of the body of work that claims that the article is not optional, but subject to a complex pattern of variation that restricts the environments in which it can be absent. They further contribute to the literature by showing that the effect of preposition has not been examined in enough depth, since we see differences in behaviour even within the set of prepositions that are generally believed to allow for the <PREP+*que*> form. As for the *en*:Definiteness interaction, I suggest that it is simply an effect of the higher acceptability of <*en*+*que*> allowing us to see the more subtle effect of Definiteness.

The data collected form the basis to the proposal through which I answer my second research question. I present a modified version of Cinque's (2020) double-headed relative clause structure. I propose that, rather than reflecting presence vs. absence, the phrase with

the article and the one without represent different underlying structures, specifically at the level of the operator phrase. The <PREP+*que*> structure contains a dP nominal phrase that is identical to the antecedent at the point of attachment; leftwards movement by the antecedent triggers the deletion of this operator phrase. The <PREP+ART+*que*> structure contains a nominal particle that is [+definite, +specific] and has a larger nominal structure than the antecedent (at least a DP). This proposal re-states the issue of article optionality in updated terms and using a contemporary framework, giving future research a concrete point of departure.

This research project was constrained by time and circumstances in terms of the kind of tasks that I could conduct. This is reflected in the fact that both the corpus and the AJT were implemented through written texts and stimuli, which might not be as reflective of everyday speakers' usage and intuition as tasks that involve speaking and listening. Furthermore, since I used a free sample of a larger corpus that had been stripped of metadata, it is possible that I glossed over demographic distinctions (such as dialect, social class, age, etc.) that would have been relevant to how I designed and implemented the AJT. It is also worth mentioning that the written nature of these tasks made them ill-suited to investigate restrictiveness of the RC as a predictive factor, since prosodic structure is the most reliable indicator of this distinction. Nonetheless, I chose to follow Porto Dapena (1997) in assuming that the <PREP+*que*> structure is only viable in restrictive relative clauses, which is reflected in the AJT stimuli and in my theoretical analysis.

My theoretical analysis is limited in that I cannot provide a satisfactory explanation for why <*en+que*> is so much more common and acceptable than <*de+que*>. This means that my results are not necessarily able to be generalized to other prepositions, let alone outside of prep-RCs or to languages other than Spanish. That said, I must also point out that the corpus did not contain a large enough quantity of <PREP+*que*> utterances to be able to do statistical analysis of any preposition other than *en*. The analytical limitations presented by the over-representation of *en* could be addressed in future research that specifically centers this issue; a corpus study that deliberately aims to have equal numbers of <PREP+(ART)+*que*> utterances for each preposition would give us a clearer understanding of the behaviour of each preposition and whether generalizations could be made. Another approach to investigating

whether we can make general claims about the effect of preposition would have been to take a much deeper dive into the study of prepositions as a category, and prepositional structure, than what I was able to do. It would have also been ideal to examine the article and its relationship with the relative operator in more granular detail.

The nominal particle that has been the focus of my research (“the article”) is very under-researched despite being “one of the most complex [phenomena] of Spanish grammar [...] subjected to a large margin of variation on a diachronic and even dialectal level” (Brucart, 1999, p. 496). I can only hope that this work has opened the doors to all the potential possibilities for the research of article optionality; I present some of the possibilities below.

In addition to the factors that I examined in the AJT and incorporated into my theoretical proposal (definiteness and preposition), the corpus also found significant effects of polarity and animacy. As I discuss in Section 5.2.1, Brucart (1992, 1994) has proposed a link between polarity and definiteness. Further research into this factor could potentially provide additional insights on the properties and mechanisms that I sketch out in my proposal. Animacy seems to be intertwined with Differential Object Marking, and its study could contribute to this complex phenomenon. There are also factors that I had to exclude early on that seem to have a clear effect on the presence/absence of the article. One factor is the subset of RCs that contain an infinitive verb and a null antecedent. Another factor is the effect of syntactic type, i.e., whether the claim that restrictiveness is a constraining factor could have empirical support. It would also be interesting to explore other types of relative clauses and whether they can be linked to the factors we have already established.

There are also specific aspects of this thesis that could be expanded upon. We could analyse specificity as a distinct factor and separate its effect from that of antecedent definiteness. This could be done through an experimental task, e.g., a contextual felicity task, that presents participants with the necessary context to be able to distinguish specific from non-specific antecedents. There is also the matter of between-participant variation. The AJT showed distinct patterns of behaviour that I did not have the statistical power to tease apart. A replication that specifically focuses on between-participant comparisons would allow us to explore the possibility of multiple populations. Examining the other prepositions that have a viable <PREP+*que*> structure, exploring whether the syllabicity constraint is best described

as such, and doing more research into prepositions in general are also important avenues of research that can elucidate whether this proposal has generalizable aspects.

REFERENCES

- Abney, S. P. (1987). *The English noun phrase in its sentential aspect*. [Doctoral dissertation, Massachusetts Institute of Technology].
- Arce, L.C. (2017). Cliticización de complementos de régimen en español. *Letralia*, 3(1), 33-47.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1-48.
- Bird, S., Loper, E., & Klein, E. (2016). *Natural language processing with python*. Retrieved from <http://www.nltk.org/book/>
- Blas Arroyo, J. L., & Vellón Lahoz, J. (2017). En los albores de un cambio lingüístico: factores condicionantes y fases en la inserción del artículo en relativas oblicuas del siglo XVIII. *Zeitschrift für romanische Philologie*, 133(2), 492-529.
- Blas Arroyo, J. L., & Vellón Lahoz, J. (2018). On the trail of grammaticalization in progress: has el que become a compound relative pronoun in the history of Spanish prepositional relative clauses? *Probus*, 30(1), 1-45.
- Brucart, J. M. (1992). Some Asymmetries in the Functioning of Relative Pronouns in Spanish. *Catalan Working Papers in Linguistics*, 113-143.
- Brucart, J.M. (1994). El funcionamiento sintáctico de los relativos en español. In A. Alonso, B. Garza & J. Pascual (Eds.), *II Encuentro de lingüistas y filólogos de España y México* (pp. 443-69). Universidad de Salamanca.
- Brucart, J. M. (1999). La estructura del sintagma nominal: Las oraciones de relativo. In I. Bosque & V. Demonte (Eds.), *Gramática Descriptiva de la Lengua Española*, Colección Nebrija y Bello (pp. 395-522). Espasa.
- Brucart, J. M. (2016). Oraciones de relativo. In J. Gutierrez-Rexach (Ed.), *Enciclopedia de Lingüística Hispánica* (pp. 722-736). Routledge.

- Cable, S. (2010a). *The grammar of Q: Q-particles, wh-movement, and pied-piping*. Oxford University Press.
- Cable, S. (2010b). Against the Existence of Pied-Piping: Evidence from Tlingit. *Linguistic Inquiry*, 41(4), 563-594.
- Cinque, G. (2020). *The syntax of relative clauses: A unified analysis*. Cambridge University Press.
- Chomsky, N. (1995). *The Minimalist Program*. MIT Press.
- Davies, M. (2004). *El uso del Corpus del Español y otros corpus para investigar la variación actual y los cambios históricos*. University Sophia.
- Déchaine, R.M. & Wiltschko, M. (2002). Decomposing pronouns. *Linguistic Inquiry*, 33(3), 409-442.
- Fábregas, A. (2013). Differential object marking in Spanish: State of the art. *Borealis—An International Journal of Hispanic Linguistics*, 2(2), 1-80.
- Fábregas, A. (2017). The syntax and semantics of nominal modifiers in Spanish: Interpretation, types and ordering facts. *Borealis: An International Journal of Hispanic Linguistics*, 6(2), 1-102.
- Grimshaw, J. (2000): Locality and Extended Projection. In P. Coopmans, M. Everaert & J. Grimshaw (Eds.), *Lexical Specification and Insertion* (pp. 115-133). John Benjamins Publishing Company.
- Guzmán Riverón, M. (2012) El artículo en las relativas oblicuas [prep. + (art. Definido) + «que»] en textos americanos del siglo XVIII. *Cuadernos dieciochistas*, 13, 175–208.
- Heck, F. (2008). *On pied-piping: Wh-movement and beyond*. Mouton de Gruyter.
- Hornstein, N., Nunes, J., & Grohmann, K. K. (2010). *Understanding minimalism*. Cambridge University Press.

- Horvath, J. (2017). Pied-piping. In M. Everaert & H. C. van Riemsdijk (Eds.), *The Wiley Blackwell Companion to Syntax, second edition* (pp. 1-64). DOI: 10.1002/9781118358733.wbsyncom064
- Lebeaux, D. (1991). Relative clauses, licensing, and the nature of the derivation. In S.D. Rothstein (Ed.) *Perspectives on Phrase Structure: Heads and Licensing* (pp. 209-239). Academic Press.
- Loock, R. (2010). *Appositive Relative Clauses in English: Discourse functions and competing structures*. John Benjamins Publishing Company.
- Lopez, A. C. (2010). *Adding and subtracting alternation: Resumption and prepositional phrase chopping in Spanish relative clauses* (Order No. 3403541). [Doctoral dissertation, University of Southern California]. ProQuest Dissertations & Theses Global.
- Kato, M. A., & Nunes, J. (2009). A uniform raising analysis for standard and nonstandard relative clauses in Brazilian Portuguese. In J. Nunes (Ed.) *Minimalist essays on Brazilian Portuguese syntax* (pp. 93-120). John Benjamins Publishing Company.
- Kayne, R. (1994). *The antisymmetry of syntax*. MIT press.
- Martínez, J.A. (1989). *El pronombre II: numerales, indefinidos y relativos*. Arco Libros.
- Moreno Fernandez, F. (2016). *La lengua española en su geografía: Manual de dialectología hispánica, tercera edición*. Arco Libros.
- Perpiñán, S. (2010). *On L2 grammar and processing: The case of oblique relative clauses and the null-prep phenomenon* (Order No. 3452103). [Doctoral dissertation, University of Illinois Urbana-Champaign]. ProQuest Dissertations & Theses Global.
- Pierce, J., Gray, J.R., Simpson, S., MacAskill, M., Hochenberger, R., Sogo, H., Kastman, E., & Lindelov, J.K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51, 195-203.
- Porto Dapena, J. Á. (1997). *Relativos e interrogativos*. Arco Libros.

- Plann, S. (1980). *Relative clauses in Spanish without overt antecedents and related constructions*. University of California Press.
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org>.
- Rivero, M. L. (1980). That-Relatives and Deletion in Comp in Spanish. *Cahiers Linguistiques d'Ottawa*, 9, 383-99.
- Rivero, M.L. (1982). Las relativas restrictivas con que. *Nueva Revista de Filología Hispánica*, 31(2), 195-234.
- Ross, J.R. (1967). *Constraints on variables in syntax*. [Doctoral dissertation, Massachusetts Institute of Technology].
- Santana Marrero, J. (2004). Preposición+(artículo)+que relativo: análisis en la norma lingüística culta panhispánica. *Boletín de Lingüística*, 21, 66-91.
- Schütze, C. T. (2011). Linguistic evidence and grammatical theory. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(2), 206-221.
- Stark, E., & Pomino, N. (2010). How the Latin neuter pronominal forms became markers of non-individuation in Spanish. In *Grammaticalization: Current views and issues* (pp. 273-294). John Benjamins
- Suñer, M. (2001). Some Thoughts on Que: Description, Theory, and L2. *Hispania*, 83(4), 867- 876.
- Ursini, F.A. (2013). On the syntax and semantics of Spanish spatial prepositions. *Borealis: An International Journal of Hispanic Linguistics*, 2(1), 117-166.
- Vellón Lahoz, F. J. (2019). El artículo en las relativas introducidas por con en el siglo XVIII: contextos y evolución. *Revista de Filología Española*, 99(2), 391-415.

- Vellón Lahoz, F. J. (2020). Las cláusulas de relativo con antecedente humano en español: la difusión de la forma el + que entre los siglos XIX y XX. *Revista del Instituto de Lengua y Cultura Españolas*, 6(2), 769-95.
- Wiltschko, M. (2012). What does it take to host a (restrictive) relative clause? *Working Papers of the Linguistics Circle*, 21(2), 100-145.
- Winter, B. (2013). *Linear models and linear mixed effects models in R with linguistic applications*. Retrieved from: <http://arxiv.org/pdf/1308.5499.pdf>
- Zulaica Hernández, I. (2018). A denotation-driven reanalysis of the Spanish neuter pronominal system. *Lingua*, 213, 91-116

APPENDIX

LIST OF TARGET STIMULI FOR THE ACCEPTABILITY JUDGEMENT TASK.

For all pairs, the argument condition is (#a) and the adjunct condition is (#b).

- (A1) a. *Los chimpancés se comieron unas/las palmeras de (las) que se acordaban en el zoológico.*
 ‘The chimps ate some/the palm trees that they remembered at the zoo.’
- b. *Los chimpancés se comieron unas/las palmeras de (las) que se colgaban en el zoológico.*
 ‘The chimps ate some/the palm trees from which they hung in the zoo.’
- (A2) a. *El perezoso destruyó unas/las ramas de (las) que se acordaba ocasionalmente.*
 ‘The sloth destroyed some/the branches that it remembered occasionally.’
- b. *El perezoso destruyó unas/las ramas de (las) que se colgaba ocasionalmente.*
 ‘The sloth destroyed some/the branches from which it hung occasionally.’
- (A3) a. *Los clavadistas admiraron unas/las plataformas de (las) que hablaban a menudo.*
 ‘The divers admired some/the platforms of which they talked often.’
- b. *Los clavadistas admiraron unas/las plataformas de (las) que saltaban a menudo.*
 ‘The divers admired some/the platforms from which they jumped often.’
- (A4) a. *El paracaidista compró unas/las avionetas de (las) que hablaba con gusto.*
 ‘The skydiver bought some/the small planes of which he talked with joy.’
- b. *El paracaidista compró unas/las avionetas de (las) que saltaba con gusto.*
 ‘The skydiver bought some/the small planes from which he jumped with joy.’
- (A5) a. *Mi tía demolió unas/las escaleras de (las) que se burlaba todo el tiempo.*
 ‘My aunt demolished some/the stairs that she mocked all the time.’
- b. *Mi tía demolió unas/las escaleras de (las) que se caía todo el tiempo.*
 ‘My aunt demolished some/the stairs from which she fell all the time.’

- (A6) a. *Los acróbatas remplazaron unas/las cuerdas de (las) que se burlaban con frecuencia.*
 ‘The acrobats replaced some/the cords that they mocked with frequency.’
- b. *Los acróbatas remplazaron unas/las cuerdas de (las) que se caían con frecuencia.*
 ‘The acrobats replaced some/the cords from which they fell with frequency.’
- (A7) a. *Mis abuelos rentaron unas/las haciendas de (las) que disfrutaban durante la semana.*
 ‘My grandparents rented some/the ranches which they enjoyed during the week(days).’
- b. *Mis abuelos rentaron unas/las haciendas de (las) que regresaban durante la semana.*
 ‘My grandparents rented some/the ranches from which they returned during the week(days).’
- (A8) a. *La turista describió unas/las playas de (las) que disfrutaba cada verano.*
 ‘The tourist described some/the beaches that she enjoyed each summer.’
- b. *La turista describió unas/las playas de (las) que regresaba cada verano.*
 ‘The tourist described some/the beach.pl from which she returned each summer.’
- (A9) a. *La joven quemó unas/las cartas en (las) que pensaba demasiado.*
 ‘The young woman burned some/the letters about which she thought too much.’
- b. *La joven quemó unas/las cartas en (las) que decía demasiado.*
 ‘The young woman burned some/the letters in which she said too much.’
- (A10) a. *Mis primos encontraron unas/las postales en (las) que pensaban mucho.*
 ‘My cousins found some/the postcards about which they thought a lot.’
- b. *Mis primos encontraron unas/las postales en (las) que decían mucho.*
 ‘My cousins found some/the postcard.pl in which they said a lot.’

- (A11) a. *La actriz promocionó unas/las revistas en (las) que confiaba abiertamente.*
‘The actress promoted some/the magazines in which she trusted openly.’
- b. *La actriz promocionó unas/las revistas en (las) que escribía abiertamente.*
‘The actress promoted some/the magazines in which she wrote openly.’
- (A12) a. *Los senadores condenaron unas/las publicaciones en (las) que confiaban secretamente.*
‘The senators condemned some/the publications in which they trusted secretly.’
- b. *Los senadores condenaron unas/las publicaciones en (las) que escribían secretamente.*
‘The senators condemned some/the publications in which they wrote secretly.’
- (A13) a. *Los gerentes defendieron unas/las compañías en (las) que creían felizmente.*
‘The managers defended some/the companies in which they believed happily.’
- b. *Los gerentes defendieron unas/las compañías en (las) que trabajaban felizmente.*
‘The managers defended some/the companies at which they worked happily.’
- (A14) a. *La señora recomendó unas/las tiendas en las que creía desde hace tiempo.*
‘The woman recommended some/the stores in which she(‘d) believed for a while.’
- b. *La señora recomendó unas/las tiendas en (las) que trabajaba desde hace tiempo.*
‘The woman recommended some/the stores in which she(‘d) worked for a while.’
- (A15) a. *El atleta criticó unas/las carreras en (las) que participaba constantemente.*
‘The athlete criticized some/the races in which he participated constantly.’
- b. *El atleta criticó unas/las carreras en (las) que perdía constantemente.*
‘The athlete criticized some/the races in which he lost constantly.’

(A16) a. *Las bailarinas boicotearon unas/las competencias en (las) que participaban frecuentemente.*

‘The dancers boycotted some/the competitions in which they participated frequently.’

b. *Las bailarinas boicotearon unas/las competencias en (las) que perdían frecuentemente.*

‘The dancers boycotted some/the competitions in which they lost frequently.’