ARTICLE OPTIONALITY IN SPANISH PREPOSITIONAL RELATIVE CLAUSES

Andrea Levinstein Rodriguez
University of Calgary

1.   Introduction

This research project is centered around 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) La   mesa   en   **(la)**   que puse   el   libro
     The.F.SG   table   on   **ART.F.SG** QUE put.1SG   the   book
     'The table in 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 absence of the article. 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 (Martinez, 1989; Porto Dapena, 1997). We can see this in the contrast between examples (1) and (2), in which the article is 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. [1]

(2) *El   edificio   ante   que nos   encontramos   es   Barroco.
     The.M.SG building   before   QUE 1REFL.1PL   found.1PL   is   Baroque.
     'The building in front of which we found ourselves is Baroque.'

(Porto Dapena, 1997, p. 24)

This project aims to answer 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 (AJT)

---

[1] It is worth acknowledging that, while my research 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 default or underlying state is beyond the scope of this paper.

that provided positive and negative evidence, respectively, of these factors. The present article focuses on the methodology and results obtained from the AJT, which showed a significant effect of preposition and antecedent definiteness on the acceptability of sentences without the article. The data collected was then integrated into the proposal through which I answer my second research question, 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 paper is as follows: Section 2 introduces the most relevant syntactic approaches to the study of Spanish prep-RCs. Section 3 presents the Acceptability Judgement Task, and Section 4 discusses how the results of this tasks have been integrated into an updated theoretical understanding of Spanish prep-RCs. Section 5 is the conclusion.

2. Background

Section 2.1 introduces the most relevant syntactic literature on the study of Spanish prep-RCs, focusing on each work's treatment of article alternation. Section 2.2 introduces a contemporary approach to prep-RC structure that will serve to update the existing analysis.
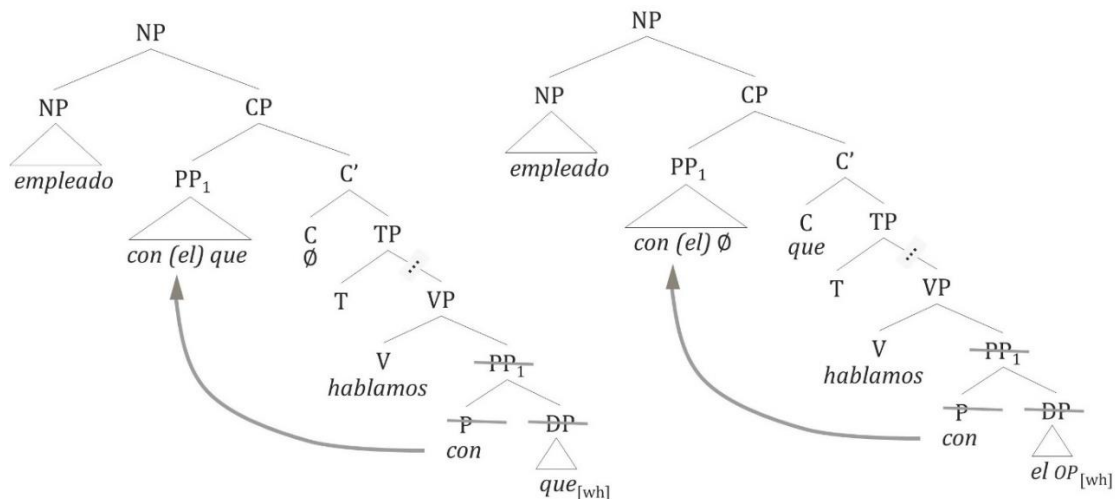
2.1 Syntactic literature on Spanish prep-RCs

Although Spanish prep-RCs themselves are not under-researched, the issue of article alternation is one that is often brushed aside or discussed only tangentially in the syntactic literature. Rather, the literature concerns itself with two topics: the distinction between prep-RCs and what Brucart (2016) calls *semilibre* relative clauses,[2] and the categorization of *que*. This section summarizes the latter issue.

 The Bare Que hypothesis (Rivero, 1980) argues that, while *que* is a complementizer in subject and object RCs, the *que* in prepositional RCs is a *wh*-word (analogous to the English *which*). Brucart's (1992, 1999) Unique Que hypothesis posits that *que* is always a complementizer (analogous to English *that*). Both approaches present a standard generativist structure for the prep-RC, with a Complementizer and an Operator that undergoes *wh*-movement. As shown in Figure 1,[3] both proposals are identical in terms of structure, with the only difference being which element is null and which one is overt. For Brucart, the Complementizer is *que* and the Operator is null; for Suñer, the Complementizer is null and the Operator is *que*. Crucially, neither proposal accounts for the article on a structural level.

---

[2] *Semilibres* are subject or object RCs that have been selected by a partially elided antecedent (Brucart, 2016); when selected by a preposition, they have an identical surface word-order to that of prep-RCs.

[3] Both trees presented here represent my best understanding of the hierarchical structures proposed by Brucart and Suñer. Brucart (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.

Figure 1. Trees for *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).

Nonetheless, both Brucart's (1992) proposal and Suñer's (2001) counterargument to it explicitly address the question of article alternation. Brucart proposes that the role of the article is to reproduce the φ-features of the antecedent, which makes the presence of the null operator (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. However, his proposal does not identify where the article might merge into the tree or what mechanism underlies its role as operator-antecedent liaison.

Brucart (1992, 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 (3) but mandatory when it contains negation (4). The second constraint is definiteness – the article is optional when the antecedent is definite (3), but mandatory when it is indefinite (5). 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 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.

(3) 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…'

(4) El   dinero   de**-l** /   *de   que no   disponía
    The.M.SG money   to**-ART.M.SG** of   QUE not had.access.3SG
       'The money to which he did not have access…'

(5)   Un   dinero   de-**l** /   *de   que disponía
       A.M.SG   money   to-**ART.M.SG**   of   QUE had.access.3SG
       'A money to which he had access…'

<div align="right">(Brucart, 1992, pp.119-20)</div>

The third constraint is better characterized as a series of observations rather than a constraint per-se. Brucart (1999) addresses the syllabicity restrictions identified throughout the literature, speculating that this effect 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.

As part of her opposition to the Unique Que Hypothesis, Suñer (2001) argues that the fact that the Unique Que hypothesis accounts for article alternation in prep-RCs is not a strong argument in its favor, because this structure is extremely rare and not reflective of everyday spoken Spanish. She uses corpus data (Bentivoglio & Sedano, 1993, as cited in Suñer, 2001) to support her claim that the <PREP+ART+*que*> structure is extremely rare in all but the most formal registers. 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).[4]

2.2   Syntactic literature on the head-antecedent relationship

As the section above shows, the research that exists to date on Spanish prep-RC does not account for article alternation at a structural level. Therefore, we turn to a more contemporary approach to the generative study of prep-RCs. Specifically, this section explores a recent proposal that re-analyses the relationship between the antecedent of a relative clause and the relative operator. Cinque (2020) expands upon a Head-Matching understanding of the head-antecedent relationship to provide a unified account of all types of relative clauses.[5] According to this analysis, all relative clauses can be derived from the

---

[4] Although this is not addressed in this paper due to space constraints, one of the goals of this wider project is to find evidence that the <PREP+(ART)+*que*> structure is part of speakers' mental grammar, regardless of its relative rarity.

[5] In a 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 (i). This can be contrasted to the Head-External Analysis, the standard approach used by Brucart (1992) and Suñer (2001), in which the operator is a single *wh*-word or a null operator (ii).

(i)   the book$_i$ [$_{CP}$ [$_{DP}$ which ~~book$_i$~~]$_1$ John likes ~~**which book₁**~~]

(ii)   the book$_i$ [$_{CP}$ [$_{DP}$ which$_i$]$_1$ John likes ~~**which₁**~~]   (Salzmann, 2018, p. 1)

same type of double-headed structure. Here, I 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 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.

Cinque (2020) incorporates an aspect of the Head-Raising analysis, 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 (6a). However, if the dP_1 moves up the nominal tree, then it licenses the deletion of dP_2 instead (6b).

(6) a. [DP the [YP [CP [dP2 two nice books]1 that John wrote [dP2 **two nice books**]1] [YP Y [dP1 two nice books]]]

   b. [DP the [dP1 two nice books]2 [YP [CP [dP2 two nice books]1 that John wrote [dP2 **two nice books**]1] [YP Y [dP1 **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. For the Spanish prep-RC, however, it suffices to look at the linear order. As (7a) 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 (7b).

(7) a. la mesa [CP [PP en la mesa]1 que puse el libro **en la mesa1**]
   b. *la mesa [CP [PP en la mesa]1 que puse el libro **en la mesa1**]

Since Cinque's (2020) double-headed structure can straight-forwardly account for the derivation of <PREP+*que*> sentences, this framework was adopted in this work as a promising avenue to explore article alternation, as will be shown in Section 4.[6]

3.   Acceptability Judgement Task

Section 3.1 summarizes the results of the corpus study (Levinstein Rodriguez, 2022) that served as the basis for the current task. Section 3.2 describes the stimuli, experimental task and participants. Section 3.3 outlines the predictions and Section 3.4 presents the results.

---

[6] Another benefit of Cinque (2020) is its compatibility with Brucart's (1992) Unique Que Hypothesis. Cinque briefly touches upon the data in Brucart (1994), broadly contrasting the Canarian Spanish data presented there with comparable structures in other languages (2020, pp. 59-62).

3.1 Background: Corpus Study

A corpus study was conducted using a free sample of the *Corpus del Español* ('Corpus of Spanish', Davies, 2004), a database of text extracted from web sources (blogs, forum posts, news articles). Using chi-square analyses, 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 the preposition used. Prepositions *en* 'in/on/at' and *de* 'of, from' were taken as representative of this propensity towards accepting vs. rejecting the absence of the article. There was also a significant 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. The preposition effect was hypothesized to reflect an underlying distinction between embedded operators in an adjunct position and those in an argument position. These results form the basis for the stimulus design of the AJT.

3.2 Methodology

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: a) the presence or absence of the article; b) the definiteness of the antecedent; and c) whether the prepositional phrase was an argument or an adjunct (henceforth, "syntactic position"). The use of prepositions *en* and *de* was counterbalanced in all conditions.

The first two variables, Article and Definiteness, are illustrated in the sentence below (8). Each variable has two levels: "Present vs. Absent" for the article, and "Definite vs. Indefinite" for the antecedent definiteness. Each sentence alternated on these two levels, forming minimal pairs.

(8) Los clavadistas admiraron **unas / las** plataformas
    the divers admired **a.F.PL the. F.PL** platforms
    de **(las)** que <u>saltaban</u> a menudo
    from ART.F.PL QUE jumped often
    'The divers admired the platforms from which they jumped often'

To examine the role of syntactic position, stimuli items alternated in terms of the verb. [7] That is, the embedded verb of items in the argument condition is one that selects a Prepositional Phrase as an argument, while verbs in the adjunct condition can have an adjunct PP. For example, (8) above, which includes the adjunct-selecting verb *saltaban*

---

[7] The argument/adjunct distinction is partially derived from the observation in Brucart (1999) that manner antecedents (associated with the adjunct position) seem to prefer the <PREP+*que*> structure; this observation is attested in subsequent corpus work. Crucially, the antecedents identified by Brucart are all preceded by the preposition *en*. For an in-depth explanation of this criterion, as well as the tests used to diagnose argument vs. adjunct antecedents, please see Levinstein Rodriguez (2022, §3.4, §4.1)

'jumped (from),' alternates with the argument-selecting verb *hablaban* 'spoke (of)' (9). The items had otherwise the same linear order.

(9)   Los   clavadistas   admiraron   **unas / las**   plataformas
       the   divers   admired   **a.F.PL the. F.PL** platforms
       de   **(las)**   que <u>hablaban</u> a menudo
       of   ART.F.PL   QUE talked   often
       'The divers admired the platforms of which they talked often'

The target stimuli consisted of 32 sets with four conditions each (article presence/absence and definiteness/indefiniteness), with the full list counterbalanced for syntactic position and preposition used. These sets were sorted into one of four lists using a Latin Square design. 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 (containing obvious subject agreement errors).

Ninety-five potential participants were recruited from the Central Mexico area to complete an online task, where they were shown the items from one of the lists and asked to rate them on a 7-point Likert scale, with 1 being "most unacceptable" and 7 "most acceptable." Of these, 79 successfully completed the study. After removing datasets that met the removal criteria, each list was trimmed to maintain Latin square balance, so that all lists had an equal number of datasets. 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).

3.3 Predictions

Participants are predicted to react to these conditions as follows: a) 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. b) Following the literature, participants are predicted to rate <PREP+*que*> sentences with a definite antecedent significantly higher than those with an indefinite antecedent. c) 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. d) If the effect of preposition reflects an 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.

3.4 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.

Table 1: Mean ratings of the target sentences in the AJT, divided by each independent variable and some relevant interactions.

Condition	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		
en	5.16	4.74	6.32	5.75	5.31
de	4.82	2.97	5.93	4.58	4.33

Results were tested for statistical significance using linear mixed effects model comparisons using the R module *lme4* (Bates et al., 2015).[8] The best-fit model, i.e., the model with the most predictive power (10), identified article presence/absence, preposition and definiteness as significantly predictive factors, as well as a random slope for Participant ID, indicating high between-participant variation.[9]

(10) 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

---

[8] Following Winter (2013), I used an additive comparison process to identify the best-fit intercept-only model. Then I added the most extensive random slope that was justified by my designed and ran a final comparison. For a step-by-step illustration of this process, as well as a thorough explanation of the reasoning behind the use of a random slope, please see Levinstein Rodriguez (2022, §4.3)

[9] Post-hoc analysis indicates that, while most participants consistently distinguished between *<en+que>* and *<de+que>* sentences, some categorically accepted or rejected the presence of the article. However, the sample size for each participant was not large enough to test for significance.

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**).

Table 2: Fixed effects for the best-fit model.

	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

Post-hoc testing consisted of Boniferroni-corrected t-tests on all subsets of the data that were deemed significant by the lmer modeling (see Table 1 for the means reported by these tests). As well as replicating the significance of the lmer model, the t-testing found that, for *<en+que>* 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**). The post-hoc testing also revealed between-participant differences.

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.

4.   Discussion

This section evaluates the results of the experimental tasks in the context of the theoretical works introduced in Section 2, specifically focusing on the definiteness as a significant

factor. [10] The aim of this section is to use said frameworks to update our existing understanding of article alternation in Spanish prep-RCs. Section 4.1 discusses how Brucart (1992) provides a potential explanation for the definiteness effect, where *<en+que>* sentences were rated significantly higher when the antecedent was definite. Section 4.2 explores in more general terms whether this distinction can be captured through Cinque's (2020) theoretical framework. Section 4.3 goes through my proposed derivations for Spanish prep-RCs with and without the article.

4.1 Grounding the definiteness data

After making the simple observation that the article is only absent when there is a definite antecedent, Brucart (1992, 1994) suggests that the definiteness effect (where a <PREP+*que*> sentence is only absent with a definite antecedent) and the polarity effect (<PREP+*que*> only occurring when there is no negation in the embedded clause) might share an underlying cause. 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.[11] 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.

  While the details and implications of polarity as a restricting factor are beyond the scope of this research project, the crucial idea presented by Brucart (1992) 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 4.3.

4.2 Updating my analysis through Cinque (2020)

As shown in Section 2.2, 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.

  By assuming that Spanish prep-RCs follow Cinque's (2020) external head structure, the optionality issue can be reframed 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 categorically identical, and that this identity (what Cinque (2020) calls "non-

---

[10] Unfortunately, an in-depth exploration of the preposition effect was beyond the scope of this research project. Some possibilities are briefly discussed in the conclusion.

[11] From context, it seems that Cinque (1990, as cited in Brucart, 1994) and Brucart (1994) use the term "non-referential" to mean "non-specific." 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 proceed with the assumption that Brucart was talking about specificity.

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 to be made lightly. 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.[12] 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*, this indicates that *cui* must be categorially distinct from the antecedent.

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, this will allow us to capture the distinction between <PREP+*que*> and <PREP+ART+*que*>.[13]
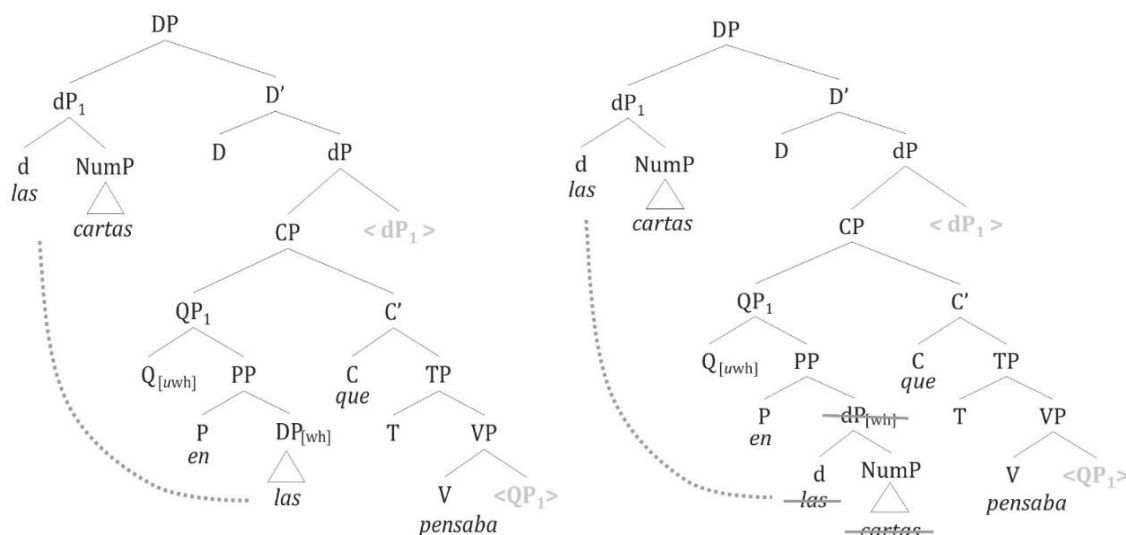
4.3   A structural account of article alternation

Cinque (2020) proposes that pronominal operators that do not undergo deletion have larger nominal structures than those that do. This fits well with Brucart'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 a segment from one of the stimulus items for an example.

---

[12] In fact, some authors have explicitly stated that there is no such difference, emphasizing that the choice to use the article in contexts that permit <PREP+*que*> must be either syntactically motivated or "purely aesthetic" (Porto Dapena, 1997, p. 24).

[13] To maintain continuity in 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 as part of the pronominal operator phrase.

Figure 2: Trees representing the <PREP+ART+*que*> structure (left) and the <PREP+*que*> structure using the sentence fragment *las cartas en (las) que pensaba*, 'the letters that she thought about'.

The left-hand side of Figure 2 represents my proposal for the <PREP+ART+*que*> structure. The operator phrase [$_{DP}$ las],[14] as well as the PP that contains it, are selected by a null Q head.[15] The QP then moves to Spec, CP. The CP modifier adjoins to an existing dP, '*las cartas*', which subsequently moves to Spec, DP. [16] 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.

By contrast, in the <PREP+*que*> structure (right-hand side), 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. This puts the external

---

[14] 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.

[15] I follow Cable's (2010) Q-Theory to account for Pied-Piping structures. This is a matter of personal preference; in principle, using a feature percolation approach should result in a mostly identical derivation.

[16] 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.

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 turn back to Brucart (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.   Conclusion

The primary goal of the project presented in this paper 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.

The primary limitation of this project is its inability to 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. It is possible that this effect is, at least partially, driven by factors external to the grammar. That is, the ubiquity of contexts that allow <*en+que*> (i.e., with the manner antecedents identified in Brucart (1999)) and disallow <*de+que*> (where *de* is part of a polysyllabic preposition, e.g., *despues de* 'after') might lead speakers to accept or reject these strings across the board regardless of whether they are technically (un)grammatical.

The corpus also 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.

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 […] subject 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 discussed in Section 4.1, 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 current 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

Bates, Douglas, Martin Maechler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67: 1–48.

Brucart, Josep Maria. 1992. Some Asymmetries in the Functioning of Relative Pronouns in Spanish. In *Catalan Working Papers in Linguistics,* 113–143.

Brucart, Josep Maria. 1994. El funcionamiento sintáctico de los relativos en español. In *II Encuentro de lingüistas y filólogos de España y México*, ed. Alegria Alonso, Beatriz Garza and Jose Pascual, 443–69. Universidad de Salamanca.

Brucart, Josep Maria. 1999. La estructura del sintagma nominal: Las oraciones de relativo. In *Gramática Descriptiva de la Lengua Española*, ed. Ignacio Bosque and Violeta Demonte, 395–522. Madrid: Espasa.

Brucart, Josep Maria. 2016. Oraciones de relativo. In *Enciclopedia de Lingüística Hispánica*, ed. Javier Gutierrez-Rexach 722–736. London: Routledge.

Cable, Seth. 2010. *The grammar of Q: Q-particles, wh-movement, and pied-piping*. Oxford: Oxford University Press.

Cinque, Guglielmo. 2020. *The syntax of relative clauses: A unified analysis*. Cambridge: Cambridge University Press.

Davies, Mark. 2004. *El uso del Corpus del Español y otros corpus para investigar la variación actual y los cambios históricos*. University Sophia.

Fábregas, Antonio. 2017. The syntax and semantics of nominal modifiers in Spanish: Interpretation, types and ordering facts. *Borealis: An International Journal of Hispanic Linguistics* 6: 1–102.

Levinstein Rodriguez, Andrea. 2022. When articles go missing: analyzing optionality in Spanish prepositional relative clauses. Master's thesis, University of Calgary. Retrieved from: http://hdl.handle.net/1880/115259

Martínez, Jose A. 1989. *El pronombre II: numerales, indefinidos y relativos*. Madrid: Arco Libros.

Porto Dapena, J. Alvaro. 1997. *Relativos e interrogativos.* Madrid: Arco Libros.

Rivero, M. Luisa. 1980. That-Relatives and Deletion in Comp in Spanish. *Cahiers Linguistiques d'Ottawa* 9: 383–99.

Salzmann, Martin. 2018. A new version of the Matching Analysis of relative clauses. In *Reconstruction Effects in Relative Clauses*. Boston: De Gruyter.

Suñer, Margarita. 2001. Some Thoughts on Que: Description, Theory, and L2. *Hispania* 83: 867–876.

Winter, Bolo. 2013. *Linear models and linear mixed effects models in R with linguistic applications.* Retrieved from: http://arxiv.org/pdf/1308.5499.pdf