Complementizers in ellipsis

A cross-linguistic study with special focus on Spanish

Max Bonke

University of Cologne

06.08.2024

Contents

Abbreviations

1 Introduction **1**

1.1 Structure of the dissertation . . . 3

2 Gapping and complementizers **7**

2.1 Gapping and complementizer optionality in German and English . . . 8

2.2 Clausal embedding in gapping theory . . . 14

2.2.1 Clausal conjuncts . . . 15

2.2.2 Sub-clausal conjuncts . . . 20

2.2.3 Variable conjunct size . . . 22

2.3 Complementizers in gapping beyond English and German . . . 28

2.4 Discourse relations in embedded gapping: Russian . . . 36

2.4.1 An empirical exploration of the coordinators *i* and *a* in locally embedded gapping and stripping in Russian . . . 39

2.4.2 Analysis and conclusions on Russian . . . 42

2.5 Factivity: Spanish . . . 43

2.5.1 Locally embedded gapping and factivity in Spanish . . . 45

2.5.2 Experimental evidence on factivity in locally embedded gapping: Bîlbîie & de la Fuente (2019) . . . 48

2.5.3 A critical discussion of Bîlbîie & de la Fuente's (2019) analysis . . . 52

2.5.4 Relevance of Bîlbîie & de la Fuente's (2019) results and analysis for globally embedded gapping . . . 57

2.6 Locally embedded gapping and factivity in other languages . . . 60

2.7 Summary and predictions for Spanish globally embedded gapping . . . 62

3 The complementizer *que* 'that' in Spanish embedded gapping **65**

3.1 Selection of the clause-embedding verbs . . . 68

3.1.1 Factivity . . . 70

3.1.2 Mood selection . . . 79

3.2 Experimental materials and design . . . 86

3.3 Results . . . 90

3.3.1 Experiment 1 . . . 90

3.3.2 Experiment 2 . . . 91

3.3.3 Comparing Experiments 1 and 2 . . . 93

3.3.4 Experiment 3 . . . 94
3.3.5 Experiment 4 . . . 95
3.3.6 Experiment 5 . . . 96
3.3.7 Experiment 6 . . . 98
3.3.8 Discussion of Experiments 1 to 6 . . . 99
3.3.9 Experiment 7 . . . 101
3.4 General discussion of all experiments . . . 104
3.5 Sketching a structure for gapping with *que* . . . 107
3.6 Conclusion . . . 116

4 An experimental study on the factivity of Spanish clause-embedding verbs 119
4.1 How to test factive presupposition quantitatively (Degen & Tonhauser 2022) . . . 120
4.2 Experimental design, materials, procedure, and participants . . . 123
4.3 Results by verbs and verb types . . . 127
4.3.1 Relating the results to claims and observations in the literature (Rivero 1971, Degen & Tonhauser 2022) . . . 132
4.4 Correlating certainty ratings and *que*-degradation in gapping . . . 135
4.5 Conclusion . . . 138

5 An exploratory approach to the meaning of *que* 'that' in Spanish embedded gapping 141
5.1 Experiment 1: Certainty . . . 143
5.2 Experiment 2: Assertion, at-issueness, and reportiveness . . . 149
5.2.1 Experimental results . . . 153
5.2.2 A critical examination of the diagnostics . . . 156
5.3 Conclusion . . . 162

6 Summary and conclusion 163

Appendix A: The complementizer *dass* 'that' in German embedded gapping 167
1 Experimental design . . . 168
2 Experiment 1: Non-factives . . . 170
3 Experiment 2: Factives . . . 171
4 Discussion and conclusion . . . 173

Appendix B: The Russian coordinators *i* and *a* in non-embedded gapping and stripping with *tože* 'too' 175

References 179

Abbreviations

1/2/3	first/second/third person
CC	propositional content of the complement clause (of a given predicate or set of predicates)
CL	clitic
DAT	dative
DOM	differential object marking
ERG	ergative
INTERROG	interrogative
IMP	imperative
IMPFV	imperfective
IND	indicative
NEG	negation
PL	plural
REFL	reflexive pronoun
SG	singular
SUBJ	subjunctive
SUP	superlative

1 Introduction

This dissertation investigates coordinated clauses that involve the ellipsis of the finite verb in the second clause. An example of such a structure is given in (1), where ellipsis has deleted the finite verb *lives* (indicated via strikethrough) in the second conjunct. This type of ellipsis, which targets the finite verb in coordinate clauses (plus potentially further constituents alongside the verb), is called "gapping" (Ross 1968). Another, related type of ellipsis is illustrated in (2), where deletion has targeted the finite verb *lives* just like in (1), plus a further constituent, the prepositional object *in Cologne*. All that is left in the second conjunct in (2) is a single constituent, the subject *Mary*, plus the polarity particle *too*. This type of ellipsis, which deletes the whole clause except for one single constituent (plus, potentially, a polarity particle), is called "stripping" (see e.g. Johnson 2018).

(1) John lives in Cologne and Mary ~~lives~~ in Berlin.

(2) John lives in Cologne and Mary ~~lives in Cologne~~, too.

The general question of this dissertation will be: what happens when gapping and stripping are embedded under a clause-embedding predicate? Examples (3) and (4) illustrate what such a structural configuration would look like. The sentence in (3) shows gapping and the sentence in (4) shows stripping; both sentences contain the clause-embedding predicate *think*. The more specific question that is crucial to this dissertation regarding these structures will be: what happens to the complementizer in these environments? For English, this question has a straightforward answer: the complementizer is ungrammatical, in both gapping, see (3), and stripping, see (4).

(3) I think that John lives in Cologne and (***that**) Mary ~~lives~~ in Berlin.

(4) I think that John lives in Cologne and (***that**) Mary ~~lives in Cologne~~, too.

The ungrammaticality of the complementizer in embedded gapping and stripping is not restricted to sentences like (3) and (4) above, but seems to be part of a more general pattern: the complementizer is not only ungrammatical when both conjuncts are jointly embedded, but also when clausal embedding obtains within the second, i.e. the elliptical, conjunct. This configuration is illustrated in (5) for gapping and (6) for stripping, again with the embedding predicate *think*. Once more, the complementizer is ungrammatical.[1]

(5) John lives in Cologne and I think (***that**) Mary ~~lives~~ in Berlin.

(6) John lives in Cologne and I think (***that**) Mary ~~lives in Cologne~~, too.

[1] There may be some discussion to be had on whether the sentences in (5) and (6) are fully acceptable without *that*. I return to this discussion in Chapter 2. What is important to the discussion at hand is that they are outright ungrammatical with *that*.

For English and some genetically or areally closely related languages such as German, Dutch or French, there is little more to be said about the complementizer in embedded gapping and stripping, other than that it is outright ungrammatical (but see Bîlbîie et al. 2023 on English and Bîlbîie et al. 2021 on French for a more nuanced view). However, there is quite a number of other, genetically diverse languages, that allow complementizers in embedded gapping and stripping. One of these languages is Spanish, which allows the complementizer *que* 'that' in both embedded gapping and stripping, and in both the structural configurations discussed for English above. The two corpus examples from the literature in (7) and (8) show embedded gapping with *que* in these two configurations. (In these examples, underscores indicate the positions of the elided verbs.)

(7) El[sic] cree que el mundo es su empresa y **que** los mexicanos ___ sus lacayos.
He thinks that the world is his company and that the Mexicans his lackeys
'He thinks that the world is his company and that the Mexicans are his lackeys.' (Bonke & Repp 2022: 528)

(8) Pero el chico la ama y dicen **que** ella ___ a él.
But the boy her loves and they.say that she DOM him
'But the boy loves her and they say that she loves him.' (Garcia-Marchena 2018: 179)

In the theoretic and experimental literature, the structure in (8), i.e. with embedding within the second conjunct, has received considerably more attention than the structure in (7). I will summarize this literature in detail in Chapter 2. What is of more relevance for now is that there are two quantitative studies on structures like (8) in Spanish specifically: Garcia-Marchena's (2018) corpus study (on stripping), and Bîlbîie & de la Fuente's (2019) acceptability rating study (on gapping). Both of these studies focus on the role of the clause-embedding verb, e.g. *dicen* '(they) say' in (8). In a nutshell, the two studies find that these structures are restricted by the type of embedding verb. With some verbs, they occur more frequently (Garcia-Marchena 2018) or show higher acceptability ratings (Bîlbîie & de la Fuente 2019) than with others. Bîlbîie & de la Fuente (2019) frame this restriction in terms of factivity: they propose that non-factive embedding verbs easily embed a gapping conjunct (as indicated by high acceptability ratings), while factive embedding verbs do not embed a gapping conjunct as easily (as indicated by degraded ratings).

In comparison to other languages, these two studies on Spanish provide a wealth of quantitative data on embedded verb ellipsis, and thus represent a valuable resource for the study of ellipsis in general. However, they focus exclusively on elliptical clauses that are embedded within their own conjunct (like in 8 above). The other structural configuration, in which the whole clausal coordination is embedded (like in 7 above), has received no quantitative treatment in the literature. This dissertation closes this empirical gap between these two structural configurations of embedded verb ellipsis in Spanish by means of an in-depth experimental investigation of the acceptability and meaning of the complementizer *que* in this specific type of embedded verb ellipsis.

A tangible need for closing this gap arises if we consider once more the relation between the two structures in English. As I point out above, in English the restrictions on the complementizer obtain in *both* structures in an equal fashion: in the elliptical clause, *that* is ungrammatical when both conjuncts are jointly embedded (see 3/4 above) and when embedding obtains within the elliptical conjunct only (see 5/6 above). The restrictions on these two structures are thus identical. If we transfer this thought to

Spanish, we may hypothesize that the restriction of the embedding verbs that Garcia-Marchena (2018) and Bîlbîie & de la Fuente (2019) observe for the one structural configuration they examine also holds in the other configuration.[2]

To test this hypothesis, I conducted an acceptability study on the complementizer in verb ellipsis (primarily gapping) in structures like (7) above. One experimental item of that acceptability study, which has the same grammatical structure as (7) above, is given in (9). The study investigated the acceptability of sentences like (9) with and without *que* in the elliptical conjunct, as a function of the type of embedding verb (i.e. *cree* 'thinks' in 9). As hypothesized, the results indicate that the acceptability of *que* is constrained by the type of embedding verb: with verbs of saying and thinking, like *cree* 'thinks' in (9), the structure is just as acceptable with *que* as without *que*. But with other verb types such as emotives, knowledge and perception verbs, desideratives or verbs with an implicit negation *que* leads to varying degrees of degradation in acceptability. These results parallel those of Bîlbîie & de la Fuente's (2019) acceptability study on embedded gapping in the other configuration, which show a similar variable degradation depending on the type of embedding verb.

(9) Santiago cree que el rey fuma cada día y (**que**) la reina, a veces.
Santiago thinks that the king smokes every day and that the queen at times
'Santiago thinks that the king smokes every day and that the queen smokes sometimes.'

The results of my acceptability study naturally raise the question of why: why is the complementizer in gapping sensitive to the type of embedding verb? My hypothesis is that *que* in gapping is indicative of a clausal structure that carries its own meaning. This meaning, by hypothesis, harmonizes with the meaning of the verbs of saying and thinking, and therefore leads to no degradation with these verbs. With other verbs, there is a soft clash of the meaning of *que* and the meaning of the embedding verb, leading to a degradation in acceptability. I explain the reasoning behind this hypothesis in detail in Chapter 3 (Section 3.5). In that chapter, I identify four candidates for what the relevant meaning may be: factivity (as was hypothesized by Bîlbîie & de la Fuente 2019 to account for their own experimental results), assertion, at-issueness and reportiveness. Verbs of saying and thinking (once more, per assumption) naturally carry these aspects of meaning in the sense that they can potentially embed an assertion, embed at-issue content, or carry a reportive function. The other verb types lack this meaning.

An experimental investigation into these aspects of meaning, however, reveals no measurable interaction of either assertion, at-issueness or reportiveness with *que* in gapping. Thus, while this dissertation provides an in-depth description of interesting acceptability patterns of the complementizer *que* in Spanish gapping, the hypothesis that these patterns relate to the meaning of clause-embedding predicates is not borne out.

1.1 Structure of the dissertation

In Chapter 2, I introduce and discuss the main topic, verb ellipsis in embedded environments, in greater detail on the basis of English and German data (Section 2.1). Section 2.2 discusses the status of embedded

[2] I derive this hypothesis from more cross-linguistic data and with a more thorough argumentation in Chapter 2

contexts in the theoretic literature on gapping and stripping. In Section 2.3, I expand the empirical range of embedded gapping and stripping beyond English and German. That section reviews the cross-linguistic literature in the search of languages that, unlike English and German, generally do allow complementizers in embedded gapping. As that section will show, there is no shortage of such languages, which indicates the ban on complementizers in English and German is not a universal feature of gapping. Furthermore, the review of the literature will show that the empirical situation regarding embedded gapping is quite poor for many languages. Section 2.4 improves on this situation in one particular language, Russian. I present the results of an acceptability judgment experiment on Russian embedded gapping and stripping, which show that this structure is not outright ungrammatical. These results cast doubt on impressionistic judgments in the theoretic literature that claim the opposite. In Section 2.5, I turn to the language with the best empirical situation when it comes to embedded gapping: Spanish. I critically review Bîlbîie & de la Fuente's (2019) acceptability study on Spanish, which will serve as the starting-off point for my own acceptability study in Chapter 3. Since the results of Bîlbîie & de la Fuente's (2019) study heavily rely on the notion of factivity, I take a comparative perspective once more in Section 2.6 to see whether the factivity patterns that Bîlbîie & de la Fuente (2019) find in Spanish are cross-linguistically robust.

From Chapter 3 onwards, the dissertation exclusively deals with Spanish (except for the appendices, see below). Chapter 3 reports the results of the acceptability study on the complementizer *que* 'that' in Spanish embedded gapping (and, to a lesser extent, stripping) that I have mentioned above. Section 3.1 discusses two crucial properties of the design of the study: factivity and mood. Sections 3.3 and 3.4 respectively present and discuss the results of the seven experiments that make up the study. In Section 3.5, I sketch a possible structural account for the results of the study. That section will also serve to derive the hypotheses as to the meaning of *que* in gapping that may account for the acceptability patterns.

Chapters 4 and 5 test the hypotheses as to the meaning of *que* in gapping that I derived in Chapter 3. Chapter 4 focusses on factivity. It reports the results of a rating experiment that collected data for a subset of the clause-embedding verbs in Chapter 3, which, in a nutshell, allow us to draw conclusions about the degree to which these verbs are factive. The results of this rating experiment show that factivity seems to be a gradual phenomenon with no abrupt differences between assumed categories of embedding verbs such as non-factive and factive. These results parallel the results of an equivalent study on English (Degen & Tonhauser 2022), after which my study was modelled, and constitute a standalone contribution to the study of factivity and the meaning of clause-embedding verbs. For the question of *que* in gapping, the data collected in Chapter 4 furthermore serves as the basis for a quantitative approach to the question whether *que* in gapping relates to factivity. On a verb-by-verb basis, there seems to be no overall meaningful relation of the factivity ratings of Chapter 4 and the acceptability data of Chapter 3, which suggests that factivity is not crucial in the acceptability of *que* in gapping.

Chapter 5 contains two experiments that test the interpretation of *que* in gapping. The first experiment applies the diagnostic of Chapter 4 directly to the gapping materials of the acceptability study (Chapter 3), to see whether the presence of *que* modulates factivity ratings. It does not. The second experiment applies innovative, exploratory diagnostics to assess whether the presence of *que* modulates the interpretation of assertion, at-issueness and/or reportiveness. Again, it does not. The overall conclusion of Chapters 4 and 5 is thus that the effect of *que* on acceptability could not be quantitatively shown to be connected to the interpretation of factivity, assertion, at-issueness, or reportiveness. The acceptability patterns of

Chapter 3 thus remain unaccounted for. Chapter 6 provides a summary and general conclusion of this dissertation.

There are two appendices to this dissertation. Appendix A contains two experiments on embedded gapping in German, which confirm the impressionistic judgments in the literature that the complementizer *dass* 'that' is ungrammatical in embedded gapping. Appendix B contains quantitative acceptability data on a detail of Russian grammar, i.e. the difference between the coordinating conjunctions *i* and *a*, which both correspond to English *and*. The relevance of these data will become apparent in the part of Chapter 2 that deals with Russian (Section 2.4). In total, this dissertation presents the results of 13 experiments on three different languages. Table 1.1 contains an overview of the experiments.[3]

Parts of this dissertation have, in accordance with the rules of the a.r.t.e.s. Graduate School for the Humanities Cologne, previously been published in two papers ("Teilpublikationen"): Bonke & Repp (2022) and Bonke (forthcoming). Bonke & Repp (2022) is on the complementizer *que* 'that' in Spanish embedded gapping. It contains parts of Chapter 2, specifically Section 2.5, and Chapter 3 (i.e. Experiments 1, 2, 3 and 6, and large parts of the description of the experimental design and materials, the discussion, and the structural analysis). Bonke (forthcoming) is on the complementizer *čto* 'that' in Russian embedded gapping. It contains parts of Chapter 2, especially Section 2.4.

I have set up an OSF directory that will host the data and scripts for the statistical analysis of the experiments in Table 1.1. The directory is available at the link in (10). At this point, the directory only contains preliminary versions of the results for all 13 experiments in the form of Excel files. They are preliminary for a number of reasons: they lack English translations of the materials, the columns in part do not have self-explanatory labels, and some of the columns contain data that is not immediately relevant to the analysis. I aim to resolve these problems until the end of this year (2024), which entails updating the results files with more useful versions, and uploading the scripts used for the statistical analyses and the generation of figures and tables. However, as they are now, the files present in the directory should go at least some way in enabling the inclined reader to comprehend and reproduce the experimental results.

(10) **Link to OSF directory containing experimental results**
https://osf.io/hycuw/

[3] Experiment 1 on Russian tested the coordinating conjunctions *i* 'and' and *a* 'and' in two different structural environments: main clauses and embedded clauses. The discussion of this experiment is split between the parts of the dissertation in which these structural environments are relevant: Chapter 2 (embedded clauses) and Appendix B (main clauses).

Table 1.1: Overview of the 13 experiments reported on in this dissertation (Experiments will be renumbered in their respective chapters, technical terms will be introduced in Chapter 2).

Experimental block	Number	Description	Chapter
Acceptability in Russian	1	Complementizer *čto* 'that' in locally embedded gapping	2
	1	Coordinators *i* 'and' and *a* 'and' in (non-embedded) gapping and stripping with *tože* 'too'	Appendix B
Acceptability of the Spanish complementizer *que* 'that' in global embedding	2	Gapping vs. stripping with *también* 'too' under indicative-taking non-factives	3
	3	Gapping vs. stripping with *también* 'too' under subjunctive-taking true factives	3
	4	Gapping with indicative-taking non-factives vs. subjunctive-taking true factives	3
	5	Full clause equivalents of Experiment 4	
	6	Gapping with indicative-taking semi-factives vs. subjunctive-taking non-factives I	3
	7	Gapping with indicative-taking semi-factives vs. subjunctive-taking non-factives II	3
	8	Gapping with all four verb types	3
Factivity in Spanish	9	The factivity of 44 Spanish clause-embedding verbs	4
Meaning of the Spanish complementizer *que* 'that' in globally embedded gapping	10	Effect of *que* on factivity	5
	11	Effect of *que* on assertion, at-issueness and reportiveness	5
Acceptability of the German complementizer *dass* 'that' in globally embedded gapping and stripping with *auch* 'too'	12	Under non-factives	Appendix A
	13	Under factives	Appendix A

2 Gapping and complementizers

In this chapter, I outline the basic facts on gapping in clausal embedding. In Section 2.1, I discuss two structural configurations that involve gapping in an embedded environment. These two structures differ as to what part of the coordinate structure is embedded: the whole coordination spanning both clauses ("global embedding", see 1), or only the clause with the elided verb in the second conjunct ("local embedding", see 2).[1] In English and German, complementizers are impossible in both these structures.

(1) John thinks that Mary practices the piano and (*that) Sarah the clarinet.

(2) Mary practices the piano and John thinks (*that) Sarah the clarinet.

While none of the basic observations on local and global embedding to be discussed in this chapter are new, the direct juxtaposition of these two structures opens up a seldom taken descriptive route on gapping. In Section 2.2 I show that existing approaches to gapping largely fail to capture the whole range of facts regarding embedded gapping, and make radically different predictions as to the acceptability of complementizers.

In Section 2.3, I extend the empirical range of this dissertation beyond English and German by looking at the distribution of complementizers in embedded gapping from a cross-linguistic angle. The core observation that I make in Section 2.3 is that there seems to be a cross-linguistic pattern as to the availability of complementizers in embedded gapping. I formulate a conjecture that says if a language allows complementizers in locally embedded gapping, it also allows complementizers in globally embedded gapping (and vice versa). I hypothesize that this conjecture generalizes not only to the general availability of complementizers in locally and globally embedded gapping, but also to subtler, language-specific constraints.

In Section 2.4, I perform a first test of this hypothesis by taking a closer look at locally and globally embedded gapping in Russian. In this language, the discourse relation between the two conjuncts constrains both embedded gapping structures in a comparable fashion. In short, when the two conjuncts are in a contrastive relation[2] with each other, the complementizer receives degraded ratings. When the conjuncts are in a parallel, non-contrastive relation, there is no such degradation. In Russian, the different discourse relations are marked by different coordinating conjunctions corresponding to English *and*: *a* for contrastive coordination, and *i* for non-contrastive coordination. I show this constraint on discourse relation by relating my experimental findings from Bonke (forthcoming) on globally embedded gapping to novel experimental evidence as well as a theoretically well-founded interpretation of the existing literature for locally embedded gapping.

Next, I turn towards the main focus of this dissertation, Spanish. In this language, locally embedded gapping is subject to constraints on the type of embedded verb. In Section 2.5, I summarize the existing

[1] I will discuss the terms "local embedding" and "global embedding" in Section 2.1.

[2] I explain what I mean by "contrastive" in Section 2.4.

literature on these verb constraints on Spanish locally embedded gapping. Specifically, I critically examine the experimental findings and analysis of an acceptability study by Bîlbîie & de la Fuente (2019). These authors argue that the crucial verb property constraining the acceptability of Spanish locally embedded gapping is factivity, such that non-factive verbs allow locally embedded gapping more easily than factive verbs. In Section 2.6, I summarize similar findings for locally embedded gapping in languages other than Spanish.

From Sections 2.3, 2.5 and 2.6, I derive predictions on globally embedded gapping in Spanish. The conjecture I make in Section 2.3 predicts that the complementizer shows the same sensitivity to the type of embedding verb, i.e. factivity, in globally embedded gapping that it does in locally embedded gapping. Bîlbîie & de la Fuente's (2019) analysis and the fact that factivity seems to constrain locally embedded gapping in a range of languages, however, predict no such effect. These divergent predictions provide the grounds for my acceptability judgment study on verb type effects in Spanish globally embedded gapping in Chapter 3.

2.1 Gapping and complementizer optionality in German and English

In German and English complement clauses, there are two conditions that each allow the complementizer to be absent. The first condition pertains to the clause-embedding predicate. Under some clause-embedding predicates, the complementizer may be omitted in the complement clause. This is illustrated in (3) for German and in (4) for English. Complementizer omission works well with some embedding predicates like German *glaubt* 'believes', see (3a), and English *thinks* in (4a). Other clause-embedding predicates however require a complementizer, such as German *bedauert* 'regrets', see the difference between (3a) and (3b), or English *regret* in (4b). When the complementizer is omitted in German, the complement clause shows the word order pattern of matrix clauses (verb second, as opposed to the verb-final order in embedded clauses with a complementizer, see 3b and 3a, respectively).

(3) a. Hans {glaubt / bedauert}, dass Maria Klavier übt.
Hans believes regrets that Maria piano practices
'Hans believes/regrets that Maria practices the piano.'

b. Hans {glaubt / *bedauert}, Maria übt Klavier.
Hans believes regrets Maria practices piano
'Hans believes/regrets that Maria practices the piano.'

(4) a. John thinks (that) Mary practices the piano.

b. John regrets *(that) Mary practices the piano.

The question of what exactly causes this optionality of the complementizer for some but not all embedding predicates has received ample attention in the literature. The relevant factors include the semantics of the embedding verb (Dor 2005, Shim & Ihsane 2017, Bîlbîie et al. 2023), its frequency (Elsness 1984, Roland et al. 2007), and how likely it is to take a clausal over a nominal complement (e.g. *regrets that Mary practices the piano* vs. *regrets his decision*, Jaeger 2010). Additionally, properties that are not tied to the embedding verb, such as the structure and interpretation of the main and/or embedded clause may affect the acceptability of complementizer omission (Bolinger 1972, Erteschik-Shir 1977, Thompson & Mulac

1991, Bošković & Lasnik 2003, Ferreira & Hudson 2011, among others). I will not discuss these factors in detail; what is relevant for the discussion at hand is the general fact that there the complementizer is optional for some, but not all clause-embedding verbs.

The second condition under which complementizers may be absent in German and English is in embedded coordinate clauses. In the second of the two coordinated clauses, the complementizer is optional, see (5) for German and (6) for English. Here, the optionality of the complementizer is independent of the embedding verb, compare the version with embedding *glaubt* 'believes' and *thinks* to their respective counterparts with *bedauert* 'regrets' and *regret* in (5) and (6).

(5) Hans {glaubt / bedauert}, dass Maria Klavier übt und (dass) Sarah Klarinette übt.
Hans believes regrets that Maria piano practices and that Sarah clarinet practices
'Hans believes that Maria practices the piano and that Sarah practices the clarinet.'

(6) John {thinks / regrets} that Mary practices the piano and (that) Sarah practices the clarinet.

As a note on terminology, I will refer to this structure, where two coordinate clauses are jointly embedded under a predicate as "global embedding". This is not a technical term from the literature; I use it here to distinguish this type of structure from a second type that is relevant to this discussion, to which I will turn below. For a brief discussion on this choice of terminology, see footnote 4 below.

Of course, global embedding potentially features two complementizers: one in the first conjunct and one in the second conjunct (see 5 and 6 above). For the present dissertation, the complementizer in the first conjunct will not be of any relevance in any regard. Only the complementizer in the second conjunct will be the subject of investigation. This means that whenever I discuss "the" complementizer in the context of global embedding, it is always the complementizer in the second conjunct that I am referring to.

One thing to note in examples (5) and (6) above is that the verbs in the two coordinated embedded clauses are identical: *übt* 'practices' in (5) and *practices* in (6). This configuration – identity of the finite verbs in coordinate clauses – licenses the ellipsis of the finite verb in the second conjunct, see examples (7b) and (8b) for an illustration in non-embedded clauses. This type of ellipsis is called "gapping" (Ross 1968). The sentences in (7a) and (8a) contain the corresponding non-elliptical clauses.

(7) a. Maria übt Klavier und Sarah übt Klarinette.
Maria practices piano and Sarah practices clarinet
'Maria practices the piano and Sarah practices the clarinet.'

b. Maria übt Klavier und Sarah ___ Klarinette.
Maria practices piano and Sarah clarinet
'Maria practices the piano and Sarah practices the clarinet.'

(8) a. Mary practices the piano and Sarah practices the clarinet.

b. Mary practices the piano and Sarah ___ the clarinet.

In examples with ellipsis, I will use an underscore to indicate the position of the elided constituents and provide non-elliptical translations like in (7b) and (8b). For the sake of consistency, I will also do this in the case of examples from the literature, even though original sources may employ different translation conventions or different means to mark the ellipsis site. I will refer to clauses with no ellipsis as "full clauses".

Returning to complementizer optionality, it has been suggested in the literature that when gapping applies in global embedding, the complementizer is no longer optional, but instead becomes ungrammatical (Fiengo 1974, Wilder 1994, 1996, 1997, Hartmann 2000, Taraldsen 2000, Repp 2009, Johnson 2018, among others). This can be seen in examples (9) and (10) for German and English, respectively.[3] These examples are the gapping equivalents of the full clauses in (5) and (6) above. In the following examples in this chapter, I will underline the embedding predicate and put the relevant complementizer (i.e. the one in the elliptical conjunct) in boldface. This will be useful for the non-German/English examples discussed from Section 2.3 onwards.

(9) Hans glaubt, dass Maria Klavier übt und (***dass**) Sarah Klarinette ___.
Hans believes that Maria piano practices and that Sarah clarinet
'Hans believes that Maria practices the piano and Sarah the clarinet.'

(10) John thinks that Mary practices the piano and (***that**) Sarah ___ the clarinet.

This pattern of the complementizer in globally embedded gapping has been noticed in the literature. Wilder (1996, 1997) formulates a constraint, the *Head Condition* in (11), and see also Lin's (2002: 33) *Chain of Deletion Generalization*, which expresses the same thought. However, the *Head Condition* does not only apply to complementizers, but covers a broader range of functional categories (e.g. auxiliaries, modal verbs, prepositions), which are not relevant for my purposes. Additionally, the focus of this section lies with descriptive generalizations rather than analyses in terms of structural properties such as c-command. I will therefore leave aside the *Head Condition* and capture the ungrammaticality of complementizers in globally embedded gapping with the observation in (12).

(11) **Head Condition on Gapping**
An ellipsis site may not be c-commanded by an overt (nondeleted) head in its domain (= conjunct). (Wilder 1997: 74)

(12) **Observation on globally embedded gapping in German and English**
The elliptical conjunct cannot contain a complementizer.

Another configuration involving clausal co- and subordination is when embedding occurs within the second conjunct, to the exclusion of the first conjunct, see (13) and (14). I will refer to this configuration as "local embedding". In this configuration, the complementizer is subject to the aforementioned constraint driven by the embedding predicate, which is why I illustrate it with embedding verbs that allow the complementizer to be absent (German *glaubt* 'believes' and English *thinks*, i.e. the same embedding verbs as in examples 3 and 4 above, which illustrate complementizer optionality in these two languages). The German example in (13a) contains the version with the complementizer *dass* and verb-final word order, and (13b) contains the version without *dass* and verb-second word order. Example (14) contains an English example with optional *that*.

(13) a. Maria übt Klavier und Hans glaubt, **dass** Sarah Klarinette übt.
Maria practices piano and Hans believes dass Sarah clarinet practices
'Maria practices the piano and Hans believes that Sarah practices the clarinet.'

[3] In Appendix A, I provide experimental evidence that the introspective judgments presented for German globally embedded gapping are accurate.

b. Maria übt Klavier und Hans glaubt, Sarah übt Klarinette.
Maria practices piano and Hans believes Sarah practices clarinet
'Maria practices the piano and Hans believes that Sarah practices the clarinet.'

(14) Mary practices the piano and John thinks (**that**) Sarah practices the clarinet.

Like the examples of global embedding in (9) and (10) above, the examples in (13) and (14) constitute environments that, in principle, license gapping: they contain coordinate clauses, and the two conjuncts contain identical finite verbs. However, when gapping applies to the variant with the complementizer, the resulting sentence is ungrammatical, see (15) and (16).

(15) * Maria übt Klavier und Hans glaubt, **dass** Sarah Klarinette ___.
Maria practices piano and Hans believes that Sarah clarinet
'Maria practices the piano and Hans believes that Sarah practices the clarinet.'

(16) * Mary practices the piano and John thinks **that** Sarah ___ the clarinet.

Locally embedded gapping features prominently in the ellipsis literature because researchers noticed the ungrammatical status of sentences like (15) and (16) early on in gapping research (e.g. Ross 1968, Sag 1977, Hankamer 1979).[4] Since then, locally embedded gapping has been used frequently as a diagnostic for syntactic structure, most prominently in the influential work by Johnson (2004 et seq). Broadly speaking, the conclusion Johnson draws from the unacceptability of (16) is that gapping cannot involve the coordination of clauses (= CPs) but instead must involve sub-clausal coordination, specifically the coordination of *v*Ps. I will return to Johnson's approach in more detail in Section 2.2. To capture the unacceptability of locally embedded gapping, Johnson (2014: 7) formulates the *No Embedding Constraint* in (17). See also Johnson's (2018) *Downward Bounding Constraint*, which goes back to an original observation by Hankamer (1979), or Boone's (2014: 11) *Equal Conjunct Requirement*, all of which (roughly) express the same thought.

(17) **No Embedding Constraint**
Let A and B be conjoined or disjoined phrases, and β be the string elided in B whose antecedent is α in A. Then α and β must contain the highest verb in A and B. (Johnson 2014: 7)

The *No Embedding Constraint* derives the ungrammaticality of (16) the following way: the two conjoined phrases are the two parts of the sentence that are connected with *and*: *Mary practices the piano* (= A) and *John thinks that Sarah the clarinet* (= B). Within B, we find that there is an elided string, i.e. the missing finite verb *practices*. This elided verb is our β. The corresponding antecedent to β in A is that part of A that is identical to β, i.e. the (non-elided) finite verb *practices* (= α). With the constituent parts to the

[4] The term "embedded gapping" is used inconsistently in the literature. Many publications (e.g. Farudi 2013, Weir 2014, Erschler 2018, Bîlbîie & de la Fuente 2019, Bîlbîie & Faghiri 2022, Bîlbîie et al. 2023 amongst others) use the term to refer to what I call "locally embedded gapping". In Bonke & Repp (2022) and Bonke (forthcoming), I use "embedded gapping" to refer to globally embedded gapping. In my opinion, this is an appropriate use of the term because gapping is defined as verb deletion in coordinate structures. Embedded gapping should therefore involve the embedding of the whole coordinate structure.

However, in this dissertation, I use the term "embedded gapping" to refer to any structure involving gapping and clausal embedding (as does Fernández-Sánchez 2021), and the terms "global" and "local" to disambiguate between the two configurations. I think this transparent nomenclature is a workable solution. The use of "locally embedded gapping" is an improvement over the inelegant term "single conjunct embedded gapping" (SCEG) used in Bonke & Repp (2022) and Bonke (forthcoming). I do not use Fernández-Sánchez's (2021) terms "Non-Canonical Gapping 1/2" because they are not transparent and I do not think the structures in question are "non-canonical".

No Embedding Constraint thus in place, we may check whether the whole example (16) satisfies the *No Embedding Constraint*. To this end, we must take a closer look at B to see whether it contains any non-elided verbs that are higher than β. The term 'higher' in this case refers to the commonly used syntactic analysis of sentence structure in the form of tree diagrams, where some elements are literally higher up the tree structure than others. For the purposes of the discussion of (16), we may substitute this term with the more intuitive term 'preceding'. Within B, we thus check whether there are any verbs preceding the elided verb β. And indeed we find such a verb preceding β: the embedding verb *thinks*. The syntactic structure of (16) is thus illicit according to the *No Embedding Constraint*. As a consequence, example (16) is ungrammatical. The ungrammaticality of the German example (15) is derived in an identical fashion.

One striking feature of the *No Embedding Constraint* is that it seems to assume that the constituent part of an embedding structure responsible for the ungrammaticality of locally embedded gapping is the embedding verb. The constituent part of an embedding structure that has been the focus of this discussion so far, the complementizer, does not play any role in the formulation of the constraint. As we will see below, this non-consideration of the complementizer constitutes a major shortcoming of the *No Embedding Constraint*. However, for the moment I will continue the discussion taking the formulation of the constraint at face value, thus ignoring the complementizer. I do so because there is more to be said about the *No Embedding Constraint* at this point.

Gapping is not the only type of ellipsis that adheres to the *No Embedding Constraint*. The other types are stripping and fragments. Stripping is illustrated in (18) and (19). Like gapping, stripping occurs in the second of coordinated clauses, but deletes the whole clause except for a single constituent (e.g. the subject *Sarah* in 18a and 18b), plus optionally a polarity particle (*auch* and *too* in 18a and 18b, respectively). Fragments are similar to stripping in that they are single clausal constituents, which remain after the deletion of the rest of the clause, see (19). However, fragments do not occur in coordinate clauses. Instead, they are frequently used as answers to questions, which is why they are often called "fragment answers" or "short answers" (Morgan 1973, Ginzburg & Sag 2001, Reich 2007).

(18) **Stripping**

a. Maria übt Klavier und Sarah ___ auch ___.
Maria practices piano and Sarah too
'Maria practices the piano and Sarah practices the piano, too.'

b. Mary practices the piano and Sarah ___ too.

(19) **Fragments/Fragment answers/Short answers**

a. A: Was übt Maria? B: Klavier.
what practices Maria piano
'A: What does Maria practice? B: The piano.'

b. A: What does Mary practice? B: The piano.

Examples (20) and (21) illustrate that locally embedded stripping and embedded fragments pattern identically to locally embedded gapping when it comes to complementizers. Stripping in local embedding is ungrammatical, see (20a) and (21a), and fragments are likewise ungrammatical in embedded clauses like in (20b) and (21b).

(20) a. * Maria übt Klavier und Hans glaubt, **dass** Sarah auch ___.
Maria practices piano and Hans believes that Sarah too
'Maria practices the piano and Hans believes that Sarah practices the piano, too.'

b. A: Was übt Maria? B: *Ich glaube, **dass** Klavier.
What practices Maria I think that piano
'A: What does Maria practice? B: I think that she practices the piano.'

(21) a. * Mary practices the piano and John thinks **that** Sarah ___ too.

b. A: What does Mary practice? B: *I think **that** the piano.

Notice furthermore that stripping like gapping does not allow complementizers in global embedding, see (22) and (23). Fragments play no role in global embedding, since they are entirely independent of clausal coordination, unlike gapping and stripping.

(22) Hans glaubt, dass Maria Klavier übt und (***dass**) Sarah auch ___.
Hans believes that Maria piano practices and that Sarah too
'Hans believes that Maria practices the piano and that Sarah practices the piano, too.'

(23) John thinks that Mary practices the piano and (***that**) Sarah ___ too.

The fact that gapping and stripping behave identically in embedded contexts suggests that these two types of ellipsis may have the same grammatical source (Johnson 2014: 8, Boone 2014: 51). Weir (2014) suggests that the embedding behaviour of gapping and fragments point to a similar grammatical source for these ellipsis types, and Reich (2007) does so, too, albeit not primarily on grounds of their embedding behaviour. The focus of this dissertation is of course gapping, but I follow these assumptions in treating gapping, stripping and fragments as exponents of the same grammatical process. Within this chapter, the difference between gapping and stripping comes up again in the discussion on Russian in Section 2.4. Stripping furthermore plays a role in Chapter 3 on Spanish globally embedded gapping.

Returning to the *No Embedding Constraint* on gapping, we note that it does not capture the whole picture of locally embedded gapping (as indicated above). A relatively recent observation has tied the presumed unacceptability of locally embedded gapping as a whole to the presence of the complementizer (see Reich 2007 for German gapping, Weir 2014 for English gapping and fragments, and Wurmbrand 2017 for stripping). In the examples above, acceptability drastically improves when the complementizer is omitted, see sentences (24) and (25) below, which are the complementizerless variants of the ungrammatical sentences (20a) and (21a) above.

(24) Maria übt Klavier und Hans glaubt, Sarah ___ Klarinette.
Maria practices piano and Hans believes Sarah clarinet
'Maria practices the piano and Hans believes that Sarah practices the clarinet.'

(25) Mary practices the piano and John thinks Sarah ___ the clarinet.

The acceptability of locally embedded gapping without the complementizer is not as straightforward as the examples above suggest. While the German example in (24) seems perfectly fine to me, the acceptability of locally embedded gapping in English has been subject to some general disagreement. I discuss this issue in more detail in Sections 2.2 and 2.6. Whatever the overall status of the acceptability of locally

embedded gapping, the relevant observation is that the complementizer leads to outright unacceptability. This is captured in the descriptive observation in (26).

(26) **Observation on locally embedded gapping in German and English**
The elliptical conjunct cannot contain a complementizer.

Comparing the two observations in (12) and (26) on globally and locally embedded gapping, respectively, we notice that the two restrictions on complementizers are identical. English and German disallow complementizers in embedded gapping generally, irrespective of whether embedding obtains globally or locally. It is therefore likely that gapping is subject to one single, general constraint, rather than two separate constraints that conspire to yield the same result. This pattern is captured by the generalization in (27), which unifies the two observations. This generalization is the core observation in this section, because it captures the relevant behaviour of embedded gapping in German and English. I assume this generalization applies to stripping and fragments (*modulo* coordination) as well.

(27) **Generalization on embedded gapping in English and German**
The elliptical conjunct cannot contain a complementizer.

The idea that there might be a unified constraint covering both local and global embedding is perhaps not surprising, given the grammatical facts outlined in this section. What is surprising, however, is that the observations leading to this generalization usually do not receive an in-depth discussion in the theoretic literature. While the facts on embedded gapping are by no means new, most contributions focus on either local or global embedding, or lack any treatment of embedded gapping altogether. In the next section, I outline the more prominent proposals of the theoretic literature on gapping. I pay special attention to the issue of embedded gapping, in both the local and global variants. I also relate the empirical observations to the overarching theoretical question of conjunct size in gapping.

2.2 Clausal embedding in gapping theory

The distribution of the complementizer impacts one of the central theoretic questions on gapping: what is the syntactic status of the conjuncts? The answer to this question is by no means obvious, given that the central element of the elliptical conjunct, the finite verb, is missing. Broadly speaking, there are two options: gapping either involves the coordination of clauses (= CPs, see 28), or the coordination of subclausal constituents (e.g. TPs, *v*Ps etc., see 29).[5] Both options have been employed in formal analyses of gapping. In this Section, I summarize these approaches, paying special attention to how they treat clausal embedding and the status of the complementizer.

(28) Mary practices the piano and [$_{\text{CP}}$ Sarah ___ the clarinet].

(29) Mary practices the piano and [$_{v\text{P/TP}}$ Sarah ___ the clarinet].

[5]The two structures in (28) and (29) are rough sketches, they do not necessarily represent any actual proposed structure from the literature.

Broadly speaking, approaches to the question of conjunct size in gapping fall into three categories: those that invariably assume clausal coordination like in (28), those that invariably assume sub-clausal coordination like (29), and those that assume variable conjunct size, i.e. (28) and (29) in alternation. From the outset, approaches involving sub-clausal coordination seem to be better suited to account for the ungrammaticality of complementizers. If the coordination obtains low enough along the clausal spine, the conjuncts simply lack the canonical position for a complementizer (= the head of CP). In globally embedded gapping, the complementizer that appears to surface in the first conjunct stands outside the coordination further up in the structure. In clausal coordination, the second conjunct contains a complementizer position. For those approaches that assume clausal coordination, the fact that the complementizer cannot surface in this position poses a challenge.

I will summarize the most prominent approaches from each of the three variants (clausal coordination, sub-clausal coordination, variable coordination) in turn. Apart from questions on embedding, I will touch upon some other issues, such as the structure of the elliptical conjunct, and the exact nature of the deletion process. The discussion will focus on English, but the empirical facts and theoretical analyses can largely be transferred to German without issues.

Before I get to all that, let me first introduce the vocabulary that is commonly used to describe the constituent parts of a gapping structure. For ease of exposition, I do so under the assumption that the elliptical conjunct is a clause, i.e. the structure in (28). The clause in the first conjunct is called the "antecedent clause" or just "antecedent" (*Mary practices the piano*). The verb in the antecedent clause, which is identical to the elided verb in the second clause, is called the "antecedent verb" (i.e. *practices*). The non-elided constituents in the elliptical clause (i.e. *Sarah* and *the clarinet*) are called "remnants" or "remnant constituents". The constituents in the antecedent clause that correspond to the remnant constituents (i.e. *Mary* and *the piano*) are called "correlates". The elliptical clause may also be referred to as the "gapped clause". It contains the missing verb, which is referred to as the "gap" or the "ellipsis site".

2.2.1 Clausal conjuncts

Clausal embedding, in both the global and local variants, has received attention in gapping research since the 1970s. However, the role of the complementizer did not feature in the discussion in many publications. Two early influential treatments of gapping,[6] Sag (1977) and Hankamer (1979), discuss local embedding, see (30) and (31), respectively. They consider their examples ungrammatical, regardless of the presence of a complementizer (contrary to the observations in Section 2.1). Hankamer (1979) also discusses globally embedded gapping, see (32), but does not mention the impact of the complementizer, instead only examining the variant without *that*.[7]

(30) * Alan went to New York, and {I know / it seems} (**that**) Betsy ___ to Boston. (Sag 1977: 139)

(31) * Alfonse stole the emeralds, and I think (**that**) Mugsy ___ the pearls. (Hankamer 1979: 19)

(32) I think that Alfonse stole the emeralds, and Mugsy ___ the pearls. (Hankamer 1979: 20)

[6] The earliest publications on gapping (under that term) are by Ross (1968, 1970). These publications do not discuss embedding.

[7] Sag (1977: 210 f.) cites Fiengo's (1974) observation on the complementizer in globally embedded gapping. I will return to Fiengo (1974) below.

The relatively low priority given to the complementizer in Sag (1977), Hankamer (1979), and early gapping research more generally is perhaps not surprising, given that the central role of the complementizer as the head of the clause was not yet a common assumption in syntactic theory at the time. The term "complementizer" was coined by Rosenbaum (1967), but the element itself was not conceptualized as the head of a clause until later (Chomsky 1986). Consequently, the contributions from that period (e.g. Ross 1968, 1970, Jackendoff 1971, Kuno 1976, Sag 1977, Hankamer 1979 and others) do not assume CP to be the syntactic category of the clause, but instead S (for "Sentence"). This category was assumed not to have a head "in any intuitive sense of that term" (Fiengo 1974: 113). It is no surprise, then, that clausal coordination was the dominant assumption in early gapping research, because under the prevailing assumptions, clausal coordination gapping did not imply the availability of the complementizer in the elliptical conjunct in embedded environments.

To account for the (presumed) ungrammaticality of locally embedded gapping, Hankamer (1979) assumes that gapping is "downward bounded [which means that] it does not go down into subordinate clauses" (Hankamer 1979: 20). This "downward bounding" is essentially equivalent to Johnson's *No Embedding Constraint*, see (17) in Section 2.1. This constraint does not only rule out locally embedded gapping, but also gapping in configurations like (33), where the elliptical clause does not connect to the antecedent via coordination, but via adverbial subordination with *while, if* etc.[8]

(33) * Alfonse stole the emeralds {because / if / although / while / unless / whenever} Mugsy ___ the pearls. (Hankamer 1979: 18)

Hankamer (1979) formulates a further constraint on gapping, which in some sense is the mirror image of "downward bounding": "upward bounding". This constraint means that if the first clause is embedded under a predicate, the second (i.e. elliptical) clause cannot be interpreted as independent (Hankamer 1979: 20). "Upward bounding" is what Hankamer (1979) illustrates with his example of globally embedded gapping in (32) above: the elliptical conjunct can only be interpreted as a complement to the embedding verb *think*, and not as forming an independent assertion. Syntactically this means that the elliptical conjunct is directly coordinated with the embedded clause in the first conjunct, see the bracketing in (34), and not with the unit of matrix and embedded clause (Hankamer 1979: 20), see (35).

[8] Some contributions in the literature claim that gapping in adverbial subordination is acceptable in certain circumstances (Izutsu 2008, Park 2016, 2019, Bîlbîie & de la Fuente 2019, Bîlbîie et al. 2023). In (i) and (ii), there are two examples that are judged to be acceptable in their respective publications. (i) is a corpus example with subordinating *before*, and (ii) is a constructed example with subordinating *while*. See Park (2019: 30 ff.) for more corpus examples with a number of different subordinating conjunctions.

(i) No doubt they will find us, before we ___ them. (corpus example from Park 2016: 300, slightly adapted)

(ii) Boys are encouraged to go out for work, while girls ___ to stay at home. (Izutsu 2008: 654)

These contributions put forward that the main condition that licenses gapping is a condition on discourse relations rather than syntax. They argue that the remnant constituents must be in a contrastive relation with their correlates in the antecedent clause for gapping to take place. This contrastive relation is most naturally obtained in coordinate clauses, which is why gapping most naturally occurs in such clauses. However, contrast may also happen in (some) subordinate clauses, which is assumed to be the case in the examples in (i) and (ii), and which licenses gapping in these examples.

Since my focus is on complement clauses and not adverbial clauses, I have nothing to contribute to this discussion. Notice however that there seems to be no quantitative acceptability data on these structures. The notion of contrast in gapping will be touched upon in this dissertation; I will return to it further below in this section and in Section 2.4 on Russian.

(34) I think [[that Alfonse stole the emeralds], and [Mugsy ___ the pearls]].

(35) * [[I think that Alfonse stole the emeralds], and [Mugsy ___ the pearls]].

One early contribution that acknowledges the relevance of the complementizer in a globally embedded environment is Fiengo (1974). He provides the examples in (36) to (38),[9] which show the expected difference in grammaticality between the variant with and without *that*, see (37) and (38), respectively.

(36) * Holmes deduced that Bartholomew smoked latakia, and **that** Thaddeus ___ trichonopoly.[10] (Fiengo 1974: 121)

(37) * Holmes deduced that Bartholomew smoked latakia, and **that** Thaddeus ___, too. (Fiengo 1974: 121)

(38) Holmes deduced that Bartholomew smoked latakia, and Thaddeus ___, too. (Fiengo 1974: 121)

Fiengo's (1974) analysis of gapping in general relies on the assumption of linear restrictions on the number of remnant constituents to the left and the right of the elided verb within its conjunct. The intuitions Fiengo (1974) adduces as evidence for this are not robust (which he acknowledges, Fiengo 1974: 116). For example, the sentence in (39) is judged to be ungrammatical because only one constituent precedes the elided verb (the subject *Max*), and two constituents follow it (the direct and prepositional objects *a wallet* and *to Harry*). Hankamer's (1979: 117) assumption is that at most one constituent may follow the ellipsis site if exactly one precedes it; however, the categorization of a sentence like (39) as ungrammatical is dubious (Johnson 2018: 595).

(39) * John gave a book to Fred, and Max ___ a wallet to Harry. (Fiengo 1974: 116)

In any case, Fiengo (1974) struggles to explain the impact of the complementizer on globally embedded stripping. The difference in grammaticality between (37) and (38) above receives no satisfactory explanation. As Fiengo (1974: 122) puts it: "I consider these data as unexplained." Sag (1977), picking up on Fiengo's (1974) observations, takes the unacceptability of the complementizer as evidence for his own formulation of the gapping rule, where at most one phrasal remnant may precede the ellipsis site. Consequently, *that* in embedded gapping is ruled out, since *that* is not a phrasal category.

The identification of constraints on gapping like Hankamer's (1979) "downward/upward bounding" or Fiengo's (1974) (putative) limit on the number of constituents is one of the main goals in early gapping research. This does not only include syntactic constraints, but also semantic, pragmatic and phonological constraints (see Fiengo 1974 and Sag 1977 for the latter, especially concerning prosodic stress, pauses and intonation). I will discuss some of the semantic and pragmatic constraints in Section 2.5, which is on locally embedded gapping in Spanish. See also Kuno (1976) for an early perspective on the semantics and usage constraints on gapping.

Another goal in early gapping research was to account for the ellipsis itself. The basic framework relied on the assumption of a specific syntactic rule that effected ellipsis. This rule was thought to interact with other syntactic rules (such as scrambling) during the derivation of the clause (Ross 1968, 1970, Sag

[9] In modern terminology, (37) and (38) constitute stripping.

[10] Latakia is a type of tobacco from the city of the same name in modern day Syria. A Trichonopoly is a type of cigar originally from the Indian city of Tiruchirappalli.

1977, Hankamer 1979). This interaction of rules was used to relate properties of gapping to word order regularities in different languages (Ross 1968, 1970, Sag 1977, Hankamer 1979). As far as I can tell, this line of research was largely unsuccessful and subsequently abandoned.[11] I will therefore not go into details.

The exact nature of the gapping rule itself differs in minor respects between the early proposals. In some, it is a deletion rule that targets the verb or a larger constituent containing the verb (Ross 1968, 1970, Sag 1977, Hankamer 1979). In others, the rule prevents the verb from receiving a phonological form in the first place (Fiengo 1974, Kuno 1976). I will not go into detail on the exact nature of the rule underlying gapping, since it is immaterial to the question of embedding and complementizers. Whatever the exact nature of the rule, the early contributions on gapping do not derive the (presumed) ungrammaticality of locally embedded gapping ("downward bounding") from the deletion rule itself or the way it interacts with other rules. Rather, the absence of embedding is a standalone condition that must be met before the gapping rule can apply.

In more recent approaches to clausal coordination gapping, the ellipsis rule receives a more specific characterization. Hartmann (2000), Féry & Hartmann (2005) and Reich (2007), amongst others, propose that gapping is a form of "radical deaccentuation" (Féry & Hartmann 2005: 69). In non-elliptical coordination like in (40), the non-identical constituents are contrastively focussed and carry pitch accents (indicated with capitalization in (40). The finite verb *hates* is identical in both conjuncts and therefore unfocussed and deaccented.

(40) SAM hates REPtiles and SANdy hates INsects. (adapted from Hartmann 2000: 162)

Under this view, gapping is a process in which the deaccentuation of the verb goes one step further. In essence, the verb (plus further elided constituents, if applicable) receives so little phonological form that they remain unpronounced altogether, see (41), the gapping equivalent of (40). The only constituents of the clause that do not undergo this deaccentuation are the remnant constituents because they carry pitch accents. This approach derives the fact that remnant constituents obligatorily form a contrast relation with their correlate in the antecedent clause.

(41) SAM hates REPtiles and SANdy ___ INsects. (adapted from Hartmann 2000: 162)

None of the approaches that make use of deaccentuation for ellipsis explicitly discuss the issue of the complementizer from this angle. Only Hartmann (2000) mentions the complementizer in the context of global embedding. She argues that gapping targets the assertion of the clause, which the complementizer hosts together with the finite verb. In Section 2.2.3, I discuss a similar approach by Repp (2009), which builds on propositional anchoring rather than assertion. Crucially, Hartmann (2000) does not explicitly discuss how her deaccentuation mechanic applies to the complementizer.

Nevertheless, one can envision how deaccentuation straightforwardly accounts for the unacceptability of *that*. In both structural configurations, the complementizer cannot participate in the contrastive focus structure to receive the necessary accent. It is obligatorily unfocussed in locally embedded gapping, since the antecedent clause features no corresponding element that it can contrast with. In globally embedded gapping, there is such an element (= the complementizer in the first clause), but since both

[11] See however Abe & Hoshi (1997) for a more recent proposal that connects properties of gapping in Japanese to head direction.

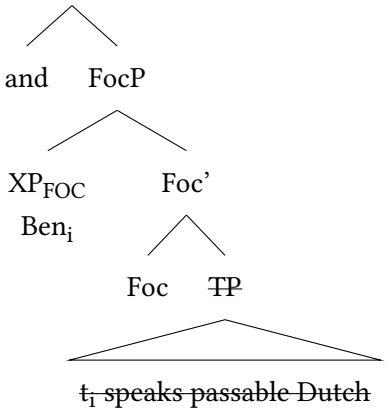
complementizers are identical, they cannot contrast with each other. Consequently, complementizers undergo obligatory deletion alongside all other deaccented constituents.

All in all, the deaccentuation approaches capitalize on the fact that remnants are prosodically prominent due to their status as focussed constituents. Another class of approaches also capitalizes on this property, focus, but from a syntactic angle. Much like the classical approaches discussed above, these approaches assume that gapping is a grammatical rule (more precisely, the featural property of a syntactic element, i.e. a head) that deletes a syntactic constituent along the clausal spine (e.g. TP). The remnant constituents originate within the deleted constituent, but move from their base positions within the TP into positions in the left periphery of the clause (= to the edge of the C-domain in structural terms) prior to deletion. These positions are outside TP and thus the constituents that occupy these positions "survive" the ellipsis.

I will illustrate one such approach, which is Wurmbrand's (2017) analysis of stripping. For the elliptical conjunct of the stripping example in (42), Wurmbrand (2017) assumes the structure in (43). The focussed remnant constituent *Ben* has moved from its position in TP to a left-peripheral position dedicated to focussed constituents (FocP).[12] TP is then deleted via a mechanism of zero-spell out, which I will not go into detail on here.

(42) Abby speaks passable Dutch, and BEN, too. (Wurmbrand 2017: 349)

(43)



(adapted from Wurmbrand 2017: 349)

Wurmbrand (2017) assumes that coordination in stripping obligatorily applies at FocP. As per Wurmbrand's (2017) assumption, this phrase is lower than the phrase that hosts the complementizer within the clausal spine (= CP),[13] which correctly rules out the complementizer in embedded stripping. Wurmbrand (2017) demonstrates this for locally embedded stripping, but the approach straightforwardly extends to globally embedded stripping, too (which she implicitly notices in her discussion of Spanish and Hungarian, Wurmbrand 2017: 360).

[12] Wurmbrand (2017), as many modern approaches to gapping and stripping, makes use of an articulated left periphery in the tradition of Rizzi (1997). This entails syntactic positions for information structural functions such as topic and focus within the C-domain. I assume that the reader is familiar with the general idea of Rizzi's (1997) proposal – the more intricate details are not relevant here. Some of the approaches on gapping and stripping discussed in the following make similar assumptions along these general lines, as do I in my analysis of globally embedded gapping in Chapter 3.5.

[13] Strictly speaking, Wurmbrand (2017) assumes sub-clausal coordination. However, since coordination in her approach obtains within the C-domain, it is more closely related to clausal coordination approaches and therefore fits into this section.

In general, this type of approach where remnants escape the elided phrase via movement is referred to as a "move-and-delete" approach. Move-and-delete approaches are popular not only for gapping, but also for other ellipsis types such as sluicing (Merchant 2001) or fragments (Boone 2014, Weir 2014). I will return to "move-and-delete" approaches in Section 2.2.3 on variable conjunct size in gapping, where I discuss another such approach by Potter et al. (2017), and also some of the general problems these approaches face. Before that, I turn to approaches that analyse gapping with (invariant) sub-clausal coordination.

2.2.2 Sub-clausal conjuncts

Approaches that assume sub-clausal coordination most prominently include the influential work by Johnson (Johnson 1996/2004, 2009, 2014, 2018), but other proposals exist, too (Coppock 2001, Lin 2002, López & Winkler 2003, Vicente 2010, Haegeman & Lohndal 2015, Toosarvandani 2016). These approaches assume coordination at *v*P or VP, from which the embedding facts (at least in part) follow naturally. Globally embedded gapping is correctly predicted to be ungrammatical without the complementizer (and only without the complementizer), since the elliptical conjunct does not contain the required structural position (= the head of CP). See the bracketing in (44), which represents a rough sketch of globally embedded gapping with *v*P-coordination.[14]

(44) John thinks [$_{CP}$ that [$_{TP}$[$_{vP}$ Mary practices the piano] and [$_{vP}$ Sarah ___ the clarinet]]]

The aforementioned proposals with sub-clausal coordination do not mention the unacceptability of *that* in globally embedded gapping. This is surprising, given that this empirical fact provides good evidence for this kind of approach. The only exception is the latest version of Johnson's work (Johnson 2018: 598), which briefly mentions the way the unacceptability of the complementizer speaks for sub-clausal coordination. In fact, Coppock (2001), Lin (2002) and López & Winkler (2003) discuss neither globally nor locally embedded gapping.[15] Toosarvandani (2016: 382) briefly mentions how the ungrammaticality of locally embedded gapping speaks for a sub-clausal coordination approach, but otherwise concentrates on deriving "Upward Bounding" within such an approach. In the following, I will focus on Johnson's work, which prominently employs locally embedded gapping as a diagnostic for the structure of the elliptical conjunct.

Johnson (2014) follows Sag's (1977) and Hankamer's (1979) judgments that locally embedded gapping is ungrammatical, regardless of the presence of *that*. See (45) for an example with the complementizer. For the version without the complementizer, Johnson (2014: 8) provides the sentence in (46), which is stripping and not gapping. Johnson (2014: 8) assumes that these two ellipsis types have the same source, so the embedding facts should be the same in both types.

(45) * Carrie will eat beans and no one knows **that** Mary ___ natto.[16] (Johnson 2014: 38)

(46) * Mary left early, and I think Sally ___ too. (Johnson 2014: 8)

[14] None of the approaches discussed in this sub-section assume the exact structure in (44), which abstracts away some of the structural details in the first conjunct (hence "rough sketch"). These structural details require some more elaboration, which I will provide below.

[15] Coppock (2001: 3) mentions the ungrammaticality of connecting the elliptical clause to the antecedent via adverbial subordination, see (33) above.

[16] Natto is a Japanese dish made with fermented soybeans (Hosking 1996: 106).

The unacceptability of locally embedded gapping, irrespective of present or absent *that*, follows from the assumption of VP- or *v*P-coordination, since such small constituents cannot host a full subordinating clause like *no one knows that* in (45) or *I think* in (46) (Johnson 2014: 38). This prediction is of course too strong because it does not differentiate between the versions with vs. without *that*. As shown earlier, the version without *that* is not outright ungrammatical.

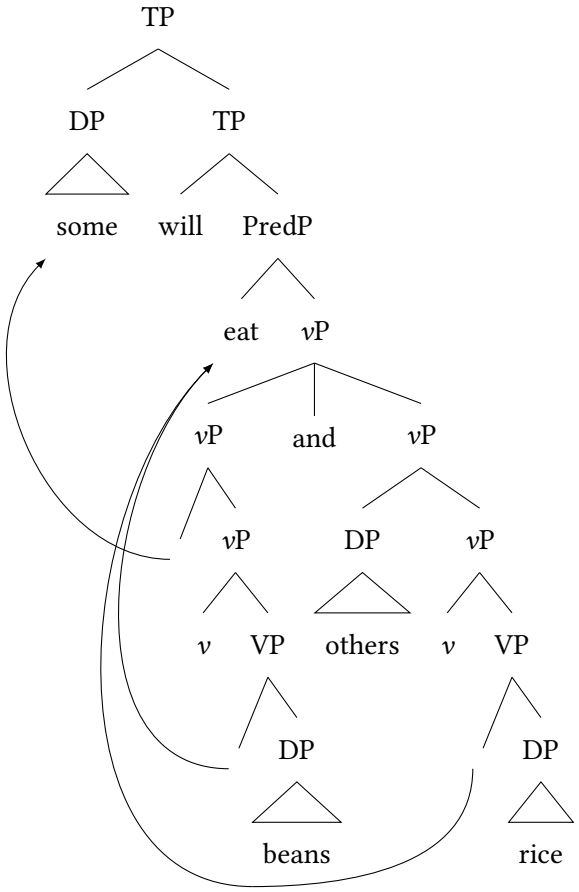
In the latest version of his analysis, Johnson (2018: 602 f.) acknowledges this effect of the complementizer in locally embedded gapping. Building on observations by Weir (2014) that in part go back to Kuno (1976), Johnson suggests that his *No Embedding Constraint*, which rules out locally embedded gapping, can be flouted under the appropriate discourse conditions. His observation, in a nutshell, is that if the second conjunct containing the embedded elliptical clause can contribute an answer to the same question as the antecedent clause, locally embedded gapping is possible. Johnson formulates this question-based discourse condition on gapping to explain why some types of embedding verbs allow locally embedded gapping more easily than others. I will go into detail on this discourse condition and the relevant verb types in Sections 2.5 and 2.6. Notice however that this discourse condition cannot account for the sensitivity to the complementizer, since the presence/absence of *that* remains unaffected by any question that is discussed in discourse. All in all, Johnson's sub-clausal coordination approach does not provide a satisfactory account for the distribution of the complementizer in locally embedded gapping. This problem is not specific to Johnson's approach, but affects all proposals that work with sub-clausal coordination (i.e. Coppock 2001, Lin 2002, López & Winkler 2003 and Toosarvandani 2016).

One aspect that sets Johnson's approach apart is the ellipsis mechanism, which relies on the movement of the verb out of the coordinated *v*Ps. For the (non-embedded) gapping example in (47), the derivation would proceed like in (48). There are two coordinated *v*Ps, which prior to movement operations are *some eat beans* and *others eat rice*. The two identical verbs then move out of their respective conjuncts simultaneously (so-called "across the board movement") into a position outside the coordination. The phrase that hosts the moved verb(s) (PredP) merges with the temporal auxiliary *will* to form a TP. The subject of the first conjunct *some* then moves to the specifier of TP for the usual case-assignment reasons.[17] For details on the more quirky aspects of this analysis (e.g. the nature of PredP or the apparent violation of the Coordinate Structure Constraint caused by the *v*-to-T movement out of the first conjunct), see Johnson (1996/2004, 2009, 2014, 2018). For criticism on Johnson's approach, see Vicente (2010) or Bîlbîie & de la Fuente (2019).

(47) Some will eat beans and others ___ rice. (Johnson 2009: 307)

[17]This is a simplification. Rather than head-movement of the verb to PredP, Johnson (2009: 307) assumes A'-movement of VP to SpecPredP, prior to which the objects in both conjuncts evacuate their thematic positions via a movement operation (Heavy NP-Shift) into an adjunct position to VP.

(48)



(adapted from Johnson 2009: 307)

Under this analysis, the fact that the finite verb appears missing in the second conjunct in gapping is not due to ellipsis in the sense that there is a specific deletion rule that targets the verb, but instead an artifact of a syntactic process (across-the-board verb movement) in the course of the derivation. The other approaches employing sub-clausal coordination mentioned in this section do not share this analysis. They either assume an ellipsis rule like that of the early proposals (Coppock 2001, Lin 2002, López & Winkler 2003) or remain agnostic as to the underlying mechanism (Toosarvandani 2016). Vicente (2010) formulates criticism of some details of Johnson's movement analysis, but does not outright reject it and offers no replacement mechanism.

I will not go into detail on the different proposals, since the differences are immaterial for the question at hand, the distribution of the complementizer. As far as this question is concerned, all the approaches involving sub-clausal coordination are on equal grounds with regard to the predictions they make: They correctly predict that globally embedded gapping is only grammatical without *that*, as mentioned above. They also correctly predict that locally embedded gapping with *that* is ungrammatical. The incorrect prediction that these approaches share is that locally embedded gapping is ungrammatical without *that*.

2.2.3 Variable conjunct size

The approaches discussed so far assume only clausal or only sub-clausal coordination. There are several alternative approaches that do not restrict the locus of coordination to only one of these options, but instead allow gapping in either clausal or sub-clausal coordination. According to these 'hybrid' approaches,

gapping is structurally ambiguous: When clauses (= TPs/CPs) are coordinated, the verb can undergo ellipsis, and if sub-clausal constituents along the clausal spine are coordinated (e.g. *v*Ps or VPs), the verb can likewise undergo ellipsis. This position is held by Repp (2009), Boone (2014), Sailor & Thoms (2014), Potter et al. (2017), amongst others.

Evidence for this structural ambiguity comes from the behaviour of scope-taking operators. The sentence in (49) contains two scope-taking operators (the modal verb *can* and a negation clitic) in the first conjunct. As observed by Siegel (1984, 1987), the interpretation of this sentence is ambiguous w.r.t. the scope of these two operators. In one reading, the conjuncts are interpreted as though each of them contained an instance of these operators (= distributive scope). The resultant reading is that of coordinated clauses containing a negated modal each, see (50). In the other reading, the operators take wide scope over the content of both conjuncts, which is illustrated by the paraphrase in (51).

(49) Ward can't eat caviar and Sue ___ beans. (Siegel 1984: 524)

(50) Ward can't eat caviar, and Sue can't eat beans. (Siegel 1984: 524)

(51) It is not possible (or desirable) for Ward to eat caviar and for Sue (simultaneously) to eat (merely) beans. (Siegel 1984: 524)

A third option is that the operators take scope within the first conjunct only (= narrow scope). This reading surfaces in more restricted contexts than the distributed and wide scope readings and is dependent on the semantic contribution of additional material in the sentence. This is illustrated in (52) and (53), which show narrow scope: the negation operator only applies within the first conjunct, the second conjunct is interpreted with positive polarity. In (52), the coordinator *but* favours an interpretation where the second conjunct differs in polarity to the first. In (53), the particle *only* has a similar effect.

(52) Pete wasn't called by Vanessa, but John ___ by Jesse. (Repp 2009: 2)

(53) Pete wasn't called by Vanessa, and John ___ only by Jesse. (Repp 2009: 2)

To account for these different scope readings, especially wide and distributive scope, the "hybrid" approaches assume different underlying syntactic structures. Examples (54) to (56) contain sketches of proposed syntactic structures for the different scopal relations: wide scope in (54), distributive scope in (55) and narrow scope in (56). Under the standard assumption that semantic scope involves an appropriate structural relation (c-command), it follows that the operators must be contained within each conjunct for distributive scope, see (55), with the subsequent ellipsis in or movement out of the second conjunct. The same is the case for narrow scope, with the difference that the second conjunct contains no operator, see (56). For the wide scope reading, it follows that the operators stand outside the coordination like in (54).

(54) [$_{TP}$Ward can't [$_{vP}$ eat caviar] and [$_{vP}$ Sue ~~eat~~ beans]].

(55) [$_{CP}$ Ward can't eat caviar] and [$_{CP}$ Sue ~~can't eat~~ beans].

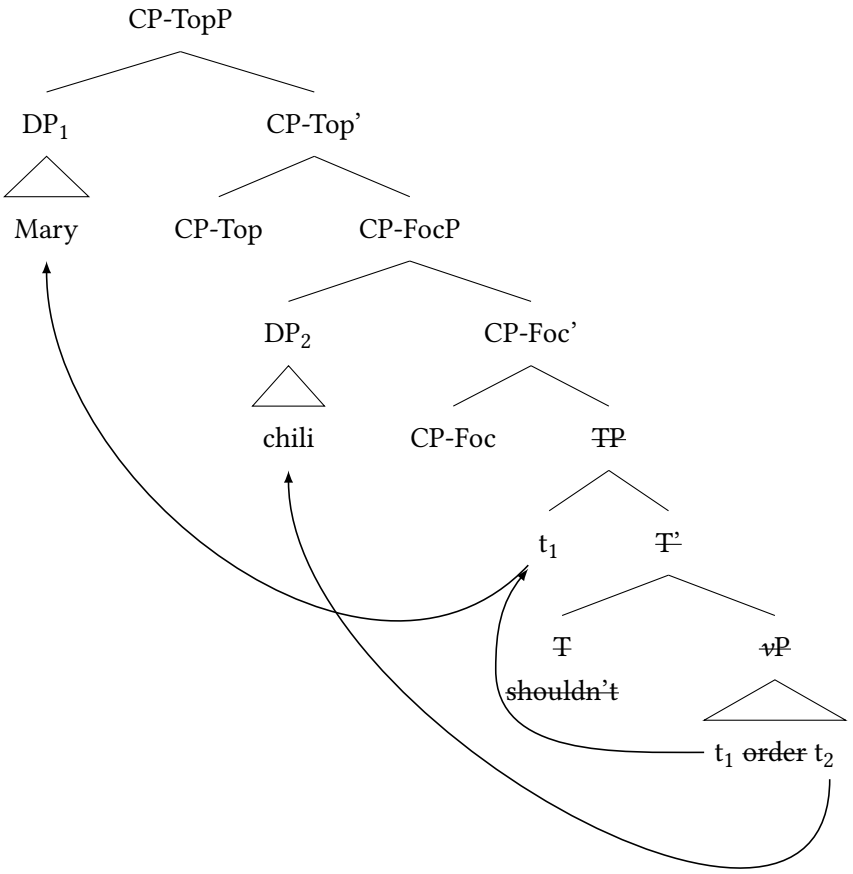
(56) [$_{CP}$ Pete wasn't called by Vanessa], but [$_{CP}$ John ~~was called~~ by Jesse].

The "hybrid" proposals make use of roughly these structures, with some differences as to structural details and the underlying ellipsis process. Boone (2014), Sailor & Thoms (2014) and Potter et al. (2017) propose

"move-and-delete" analyses (see Section 2.2.1), where ellipsis targets a phrase containing the finite verb, and remnant constituents escape the ellipsis site prior to deletion. I illustrate this for Potter et al.'s (2017) analysis of the distributed scope reading of the sentence in (57). The distributive scope reading entails the coordination of CPs, with the modal verb and negation *shouldn't* contained within the deletion site alongside the verb *order*. The structure for the crucial second conjunct is in (58). The first conjunct contains regular clause derivation and is of no interest for the discussion at hand.

(57) James shouldn't order caviar and Mary chili ___. (Potter et al. 2017: 1139)

(58)



(elliptical conjunct for the distributed scope reading of 57, slightly simplified from Potter et al. 2017: 1139)

Potter et al. (2017) assume that ellipsis targets TP, which is indicated in (58) with strike-through. Prior to the deletion, the remnant constituents *Mary* and *chili* move from the TP into positions at the edge of the C-domain. These positions are outside TP and thus are unaffected by ellipsis. Potter et al. (2017) assume that landing sites for the remnants are positions related to the information-structural notions focus (CP-Foc) and topic (CP-Top). Their proposal for gapping with sub-clausal coordination is largely equivalent, with the difference that movement of the remnants targets position at the edge of *v*P, and subsequent deletion of *v*P (Potter et al. 2017: 1139).

Potter et al.'s (2017) approach, as "move-and-delete" approaches generally, faces some challenges, one of which is that the output of the movement operations yield ungrammatical results. If no ellipsis takes

place in (58), the derivation results in the sentence in (59), which is ungrammatical. For further criticism of move-and-delete approaches to ellipsis, see e.g., Ott & Struckmeier (2018) on fragments.

(59) * James shouldn't order caviar and Mary chili shouldn't order.

In their discussion of this movement of the remnants, Potter et al. (2017) touch upon gapping in embedded clauses. They follow a claim by Hooper & Thompson (1973) that fronting into the edge of CP is impossible in the argument clauses of certain clause-embedding predicates. One of these predicates is *(be) true*. In the sentence in (60), *be true* embeds a subject clause with the regular, unmarked SVO word order. Potter et al. (2017: 1156 f.) claim that in the argument clause, the object *the soup* cannot undergo fronting into a dislocated position before the subject, as illustrated in (61). This fronting would proceed unimpeded under different embedding verbs.

(60) That James can afford the soup is true.

(61) * That the soup, James can afford is true. (Potter et al. 2017: 1157)

Due to this ban on fronting, Potter et al. (2017) argue that if the subject clause of *be true* contains gapping, only sub-clausal coordination is possible. As a result, operators can only be interpreted as taking wide scope. An example of globally embedded gapping under *be true* is in (62), where according to Potter et al. (2017: 1157), the modal verb *can* obligatorily receives wide scope. The distributive scope that would result from remnant fronting and clausal coordination is claimed to be blocked. As evidence for this blocking, Potter et al. (2017) argue that in the context in (63), where the wide scope interpretation is true and the distributive interpretation is false, the sentence in (62) is false.

(62) That James can afford the soup and Mary the sandwich ___ is true. (Potter et al. 2017: 1157)

(63) Context: James and Mary together have only $10, which is enough money for one soup or enough money for one sandwich, but not both. (Potter et al. 2017: 1157)

This argument contains two substantial empirical weaknesses: one, the ungrammaticality of fronting under *be true* seems questionable.[18] Two, the difference between the two scopal relations within the true-false paradigm of (62) and (63) might be too subtle to allow for categorical judgments. But even if we ignore these two weaknesses and concentrate on the issue of the complementizer, Potter et al.'s (2017) approach makes wrong predictions. They do not discuss those embedding predicates that do allow focus movement in their complement clause. One such predicate, according to the classification by Hooper & Thompson (1973), is *appear*. As shown in (64), the direct object *this book* within the complement clause of *appear* may be fronted to the edge of the CP.

(64) It appears that [this book]$_i$ he read t_i thoroughly. (Hooper & Thompson 1973: 478. I added the bracketing, indices, and trace.)

[18] I thank David Pesetsky for discussing the issue of fronting in embedded clauses with me. While we did not discuss Potter et al.'s (2017) specific example (61), fronting does not seem to be outright ungrammatical in complement clauses of specific embedding verbs generally, as is sometimes claimed in other publications. In any case, a claim that is presented with little to no supporting evidence, such as Potter et al.'s (2017) claim as to the ungrammaticality of (61), may be questioned or outright rejected with little to no supporting evidence as well.

If we substitute *appears* for *is true* in example (62) above, clausal coordination gapping should thus be possible. This does not only mean that a distributed reading of the modal should be available, but also that the elliptical clause should host a complementizer. Depending on which of the two CPs in structure (58) above Potter et al. (2017) assume to host the complementizer, two potential structures for globally embedded gapping emerge. Either *that* is in CP-Top, thus following the first remnant constituent, or *that* is in CP-Foc, following the second remnant. Both of these options are ungrammatical, see (65) and (66), respectively. If we assume that focus movement targets a position below the complementizer like in (61) and (64) above, the result is still ungrammatical, see (67).

(65) * It appears that James can't afford the soup and Mary **that** the sandwich ___.

(66) * It appears that James can't afford the soup and Mary the sandwich **that** ___.

(67) * It appears that James can't afford the soup and **that** Mary the sandwich ___.

In sum, the "hybrid" approach by Potter et al. (2017) does not make the correct predictions with regard to globally embedded gapping. Further, none of the "hybrid" approaches discusses locally embedded gapping at all. These problems obtain in addition to the general problems of the "move-and-delete" approach mentioned above. The other approaches I mentioned earlier that also make use of "move-and-delete" analyses (Boone 2014, Sailor & Thoms 2014) share these problems.

Within the "hybrid" analyses, Repp (2009) takes a different approach. In her analysis, the derivation of the elliptical conjunct proceeds with "sideward movement" (Nunes 2001, 2004), which is different from the fronting movement of the "move-and-delete" approaches. "Sideward movement" means that the derivation proceeds normally in the elliptical conjunct, with the special situation that the verb is not retrieved from the lexicon, but instead copied from the first conjunct, whose derivation is already complete. This copy-and-merge operation is in its mechanics identical to the implementation of classical movement operations within Minimalist syntax. However, it differs from "regular" movement because a syntactic object is not copied from a lower position in the syntactic structure and merged again in a higher position (like the fronting of the "move-and-delete" approaches, see e.g. the tree structure in 58 above), but instead movement proceeds between conjuncts. As a result, there is no lower and higher copies, but instead two structurally parallel ones, which is why this type of movement is called "sideward movement". Crucially for gapping, the copied verb is only spelled out in the first conjunct due to chain-reduction mechanisms. Thus, similar to Johnson's approach, there is no dedicated ellipsis rule in Repp (2009) and gapping is deletion in appearance only, due to the interaction of sideward movement and chain reduction during the derivation of the clause.

Repp (2009) discusses the ungrammaticality of the complementizer in globally embedded gapping. She provides the example in (68).

(68) I believe that Peter will travel with his wife to India and (***that**) Martin ___ with his colleagues to Switzerland. (Repp 2009: 210, brackets and labels omitted.)

Repp's (2009) analysis has the power to account for the absence of *that* in (68). Like the verb, the complementizer may undergo sideward movement and surface in the first conjunct only. This means that even though the elliptical conjunct is a CP with the normal complementizer position, the complementizer does

not enter the elliptical conjunct via the numeration, but via copy and merge of the complementizer in the first conjunct. Chain reduction then dictates that the complementizer is spelled out in the first conjunct only.

However, the possibility of sideward movement of the complementizer does not account for the fact that its absence from the elliptical conjunct is obligatory. This is to say, while sideward movement allows the complementizer to be missing, it does not guarantee it. To remedy this, Repp (2009) proposes a semantic-pragmatic generalization about gapping. This generalization states that gapping is the "cop[ying of] the anchoring of the proposition to the factual world" (Repp 2009: 207). Rather than targeting the finite verb specifically, then, gapping targets all elements whose semantics contribute to the relation between the proposition within the gapping conjunct, and the world in which it is evaluated. (See Section 2.2.1, where I discuss a similar approach by Hartmann 2000, which builds on assertion rather than anchoring.) Repp (2009) explains how under this generalization, the copying mechanism in gapping obligatorily applies to a number of functional elements within the clause, including complementizers. I will not go into details. Even though Repp (2009) does not discuss locally embedded gapping, her framework might provide an explanation for the obligatory absence of the complementizer in locally embedded gapping: the (non-embedded) antecedent clause contains no complementizer that might be the target for copy and merge mechanisms.

In sum, most of the approaches presented in this Section fail to account for the generalization on embedded gapping from Section 2.1 above, repeated here in (69). This is in large parts due to the fact that little attention is paid to the embedding facts generally. Contributions that do discuss embedded gapping often concentrate on globally or locally embedded gapping, but not both. Additionally, incorrect claims as to the ungrammaticality of locally embedded gapping without *that* early on in gapping research impeded adequate description. As a consequence, most of these approaches make incorrect predictions as to the embedding facts. However, I have pointed out some approaches that, even though they do not discuss both configurations in appropriate depth, could plausibly be extended to encompass the facts on both local and global embedding. These are the "move-and-delete" approach by Wurmbrand (2017), the radical deaccentuation approach by Hartmann (2000) and others, and the "sideward movement" approach by Repp (2009). An elaborated theoretical proposal along the lines of one of these approaches is beyond the scope of this dissertation. In the following section, I will take a step back from the theoretic angle of this section. Instead, I take a closer look the status of the generalization in (69) from a cross-linguistic perspective.

(69) **Generalization on embedded gapping in English and German**
The elliptical conjunct cannot contain a complementizer.

This means I will examine the available literature on both locally and globally embedded gapping for languages beyond English and German to see whether they exhibit similar patterns. As we will see, the generalization in (69) does not apply to all languages because some languages allow complementizers in embedded gapping. Nevertheless there seems to be a cross-linguistic pattern that constrains complementizers in embedded gapping. This cross-linguistic approach will aid in formulating hypotheses for the main focus of this dissertation, Spanish.

2.3 Complementizers in gapping beyond English and German

To set out my cross-linguistic investigation of complementizers in embedded gapping, let me make the (naive) conjecture that what we observe in English and German holds universally. That is, I assume that no language in the world is such that it allows complementizers in either locally or globally embedded gapping. This assumption is formulated as a conjecture in (70). I call this "Version I" of the conjecture because it will undergo some revisions to account for the cross-linguistic data presented in this section.

(70) **Typological Conjecture on embedded gapping (Version I, to be revised)**
The elliptical conjunct cannot contain a complementizer.

Version I of the Typological Conjecture gets us about as far as you would expect. It neatly captures some other (closely related) languages, see for instance the examples for Dutch globally and locally embedded gapping in (71) and (72), respectively. Both structures are only grammatical without the complementizer *dat* 'that'.

(71) **Dutch, globally embedded gapping**
Jan vertelde dat Els ziek is en (***dat**) Marie afwezig ___.
Jan told that Els ill is and that Marie absent
'Jan said that Els is ill and Marie absent.' (Broekhuis & Corver 2019: 305)

(72) **Dutch, locally embedded gapping**
Jan leest het artikel en ik denk (***dat**) Marie het boek ___.
Jan reads the article and I think that Marie the book
'Jan is reading the article and I think Marie is reading the book.' (Broekhuis & Corver 2019: 287)

We also find a similar pattern for French: globally embedded gapping is only grammatical without the complementizer, see (73). Locally embedded gapping is likewise ungrammatical, as indicated by the example in (73). See, however, the acceptability study in Bîlbîie et al. (2021), which seems to suggest that the acceptability of locally embedded gapping in French might be more subtle than the star in (74) suggests. (Bîlbîie et al. 2021 is in French. I can't provide any more specific details on their study because I don't speak French.)

(73) **French, globally embedded gapping**
Je crois que Jean voyagera à Strasbourg et (***que**) Maria ___ à Paris.
I believe that Jean travel.FUT to Strasbourg and that Maria to Paris
'I believe that Jean is travelling to Strasbourg and that Mary is travelling to Paris.' (Repp 2009: 226)

(74) **French, locally embedded gapping**
* Paul a mange une pomme et on m'a dit **que** Marie ___ une orange.
Paul has eaten an apple and one me-has told that Marie an orange
'Paul ate an apple and one told me that Marie ate an orange.' (Abeillé et al. 2014: 233)

Notice that in (74), the ungrammaticality is marked for the whole sentence, rather than for just the complementizer. French lacks the general option of omitting complementizers in simple (i.e. non-coordinated) complement clauses, unlike English, German and Dutch (see Section 2.1 for German and English). I assume that it is still the complementizer that is responsible for the ungrammaticality in (74), which then

"projects" to the whole phrase due to the lack of a grammatical variant without a complementizer. The Typological Conjecture is specific to gapping and thus not concerned with language-specific properties that are independent of verb ellipsis such as complementizer omission. These assumptions on French syntax and the Typological Conjecture allow me to conclude that French conforms to the conjecture.

The situation in French illustrates a general issue with the Typological Conjecture: if, in a given language, the omission of complementizers is generally not available in simple complement clauses, the specific contribution of the complementizer to the acceptability of locally embedded gapping will be hardly distinguishable from other potential influencing factors (whatever these may be). That is, if we find that locally embedded gapping is ungrammatical in such a language, we will not be able to test whether it is indeed the complementizer that causes this ungrammaticality in the same way that we tested it for English, German and Dutch, where leaving out the complementizer yielded grammatical sentences. One way of dealing with this issue is to just assume that the complementizer causes the ungrammaticality of locally embedded gapping, as I just did for French. In the absence of any applicable tests for the specific contribution of the complementizer, this assumption seems justifiable. The rest of this section is mostly concerned with languages that actually do allow complementizers in locally embedded gapping. This issue of complementizer omission therefore will not be relevant for these languages. However, this issue will play a role again in the discussions on Russian (Section 2.4) and Spanish (Section 2.5), where I will make similar assumptions as for French.

A look beyond the small set of genetically and areally closely related languages that I have hitherto discussed reveals that the first version of the conjecture is not accurate. In Farsi, the complementizer *ke* can surface in locally embedded gapping (Farudi 2013, Bîlbîie & Faghiri 2022), see (75). The complementizer in Farsi locally embedded gapping is, on the whole, optional. In an acceptability study, Bîlbîie & Faghiri (2022) report that the version without *ke* is ever so slightly preferred to the version with *ke* (mean ratings of 8.15 vs. 7.84, respectively, on a 0-10 scale). This difference is statistically significant (Bîlbîie & Faghiri 2022: 576 f.).

(75) **Farsi, locally embedded gapping**
Jiān be Sārā gol dād va fekr mi-kon-am **ke** Ārtur be Giti ketāb ___.
Jian to Sarah flower gave.3SG and think IMPFV-do-1SG that Arthur to Giti book
'Jian gave flowers to Sarah and I think that Arthur gave books to Giti.' (Farudi 2013: 76)

In Hindi, the complementizer *ki* 'that' is likewise reported to be acceptable in locally embedded gapping. Farudi (2013: 94 f.) provides a number of examples, one of which is in (76). However, there seems to be some nuance to the acceptability of Hindi locally embedded gapping: Kush (2016) reports that he was "unable to corroborate the judgments reported in Farudi (2013) with [his] own informants" (Kush 2016: 282), and that while some informants accept locally embedded gapping, they are a minority. For locally embedded stripping, Messick & Alok (2021: 250) report that *ki* is optional, but slightly preferred, see (77). Thus, while local embedding with the complementizer *ki* seems to be possible in Hindi gapping and/or stripping in some way, the exact extent of the phenomenon requires further empirical research. In any case, the data in (76) and (77) strongly suggest that the current version of the Typological Conjecture is false.

(76) **Hindi, locally embedded gapping**
Ram Sita=ko kitaab de-gaa aur mujhe lagtaa hai **ki** Mahesh ___ Rina=ko.
Ram Sita=DAT book give-FUT and me seem is that Mahesh Rina-DAT
'Ram will give a book to Sita and I think that that Mahesh will give a book to Rina.' (Farudi 2013: 94, glosses simplified)

(77) **Hindi, locally embedded stripping**
Sita-ne daawaa kiyaa ki Ram use bahar ghumaane le jaa saktaa hai, lekin vah nahii sochtii
Sita-ERG claim do ki Ram her out visit take go can be, but she NEG think
(**ki**) Mohan ___ bhii.
that Mohan also
'Sita claimed that Ram would ask her out, but she didn't think that Mohan would ask her out, too.' (Messick & Alok 2021: 250, glosses simplified)

Outside the Indo-European family, Shingazidja[19] allows the complementizer *uká(ya)* 'that' in locally embedded gapping, see (78).

(78) **Shingazidja, locally embedding gapping**
Djumwá hwanza ndróvi ngam-fikiri (**uká(ya)**) Saidí ___ yembe.
Juma like bananas I-think that Saidi mangoes
'Juma likes bananas, and I think that likes Saidi mangoes.' (Cédric Patin/Kassim Mohamed-Soyir, p.c. Based on Patin & Manus 2018: 920)

Similarly, Hungarian allows the complementizer *hogy* 'that' in locally embedded stripping, see (79). The analysis put forward by van Craenenbroeck & Lipták (2006, 2008, 2013) treats examples like (79) as instances of another type of ellipsis (sluicing). For our purposes, however, the example in (79) is stripping: it is a coordinate structure, with the ellipsis of the second clause save for a single constituent (the object *Bélát*).

(79) **Hungarian, locally embedded stripping**
János kirugott valakit, és azt hiszem **hogy** ___ Bélát.
János fired someone-ACC and that think-1SG that Béla-ACC
'János fired someone and I think that János fired Béla'. (van Craenenbroeck & Lipták 2013: 514, slightly adapted)

The examples from Farsi, Hindi, Shingazidja, and Hungarian show that the Typological Conjecture cannot be accurate. All members in this group of genetically diverse languages allow complementizers in locally embedded gapping and/or stripping. However, local embedding is only one of the two configurations of interest. I may thus hypothesize that the Typological Conjecture holds for globally embedded gapping. As the data from the following languages shows, this is not the case.

Georgian allows the complementizer *rom* 'that' in both configurations of interest, see (80) for local embedding, and (81) for global embedding. In global embedding, the complementizer seems to be optional. My informant[20] indicated a preference for the omission of the complementizer, which is why the

[19] Shingazidja is a Bantu language native to Grande Comore, the largest island of the Comoros (Patin & Manus 2018). The Comoros are an independent island nation in the Indian Ocean roughly between Madagascar and the northern coast of Mozambique.

[20] I thank Nino Burdiladze for discussing the Georgian data with me.

complementizer is marked with a '?' in (81). According to my informant, there is no such issue for the complementizer in the local embedding example (80), which replicates the judgment in Erschler (2018). More, ideally quantitative, research into embedded gapping in Georgian is needed to see whether there is a difference between the two structures.

(80) **Georgian, locally embedded gapping**
ia svams čais da vpikrob **rom** uča ___ ɣwinos
Ia drinks tea and I.think that Ucha wine
'Ia drinks tea and I think that Ucha drinks wine.' (Erschler 2018: 4)

(81) **Georgian, globally embedded gapping**
vpikrob, rom ia svams čais, da (?**rom**) uča ___ ɣwinos
I.think that Ia drinks tea and that Ucha wine
'I think that Ia drinks tea and that Ucha drinks wine.'

Turning back to the Indo-European languages of Europe, we observe that the sensitivity to complementizers in German, English, Dutch and French does not seem to be a genetic or areal property. Like Georgian, Romanian allows the complementizer *că* 'that' in both configurations, see (82) for local, and (83) for global embedding. There is, however, disagreement on the acceptability of locally embedded gapping in the literature: Hoyt & Teodorescu (2012: 98) report that locally embedded gapping is ungrammatical in Romanian. In fact, they report that the sentence in Abeillé et al.'s (2014) example in (83) is only acceptable if the first conjunct is embedded as well. Quantitative acceptability data in Bîlbîie et al. (2021) however seems to suggest that locally embedded gapping is indeed acceptable in Romanian.

(82) **Romanian, globally embedded gapping**
Cred că Andrei a luat cartea şi **că** Marga ___ atlasul.
I.believe that Andrei has taken the.book and that Marga the.atlas
'I believe that Andrei took the book and that Marga the atlas.' (Hoyt & Teodorescu 2012: 98)

(83) **Romanian, locally embedded gapping**
Andrei a luat cartea şi cred **că** Marga ___ atlasul.
Andrei has taken the.book and I.think that Marga the.atlas
'Andrei took the book and I think that Marga took the atlas.' (Abeillé et al. 2014: 233)

Spanish will be the main focus of this dissertation, especially in regard to globally embedded gapping. In the present section, I will briefly mention Fernández-Sánchez (2021), which is the only contribution in the literature (as far as I am aware) to provide introspective judgments on both configurations. Fernández-Sánchez (2021) reports a mismatch in acceptability between the two configurations: locally embedded gapping with the complementizer *que* is grammatical, see (84). Globally embedded gapping however is judged to be ungrammatical with *que* 'that', see (85) and (86). Notably, this ungrammaticality is a specific effect of the complementizer: without *que* these structures are fine.

(84) **Spanish, locally embedded gapping**
Alfonso robó las esmeraldas y {creo / imagino / …} **que** Mugsy ___ las perlas.
Alfonso stole the emeralds and think imagine that Mugsy the pearls
'Alfonso stole the emeralds and I think/imagine that Mugsy stole the pearls.' (Fernández-Sánchez 2021: 119)

(85) **Spanish, globally embedded gapping**
Supongo que María traerá las bebidas y (***que**) Pedro ___ el postre.
suppose that María will.bring the drinks and that Pedro the dessert
'I suppose that Mary will bring the drinks and that Pedro will bring the dessert.' (Fernández-Sánchez 2021: 122 f.)

(86) **Spanish, globally embedded gapping**
Me desagrada que Pedro me haya servido la sopa fría y (***que**) su mujer ___ el helado derretido.
me displeases that Pedro me has served the soup cold and that his wife the ice-cream melted
'It displeases me that Pedro has served me the soup cold and that his wife has served me the ice-cream melted.' (Fernández-Sánchez 2019: 121, 123)

While Fernández-Sánchez's (2021) judgments on globally embedded gapping may be accurate for the specific constructed sentences in (84) to (86), there is ample reason to believe that they are not representative for the globally embedded gapping as such. In Chapter 3, I will present quantitative acceptability data that shows that *que* is acceptable in globally embedded gapping. For now, suffice it to say that *que* cannot be ungrammatical in globally embedded gapping because it features in naturally occurring sentences, see the corpus example in (87). I will present more corpus examples of the same sort in Section 2.5.

(87) **Spanish, globally embedded gapping**
El[sic] cree que el mundo es su empresa y **que** los mexicanos ___ sus lacayos.
He thinks that the world is his company and that the Mexicans his lackeys
'He thinks that the world is his company and that the Mexicans are his lackeys.' (corpus example from Bonke & Repp 2022: 528)

Fernández-Sánchez's (2021) judgment on locally embedded gapping seems to be accurate. Actual usage examples of this structure can be found in corpora, see the example in (88). The literature features some more in-depth discussion on locally embedded gapping in Spanish (especially Bîlbîie & de la Fuente's (2019) acceptability study), which I will discuss in detail in Section 2.5. For now, I will conclude that Spanish allows the complementizer *que* in both locally and globally embedded gapping (in the latter aspect contra Fernández-Sánchez 2021).

(88) **Spanish, locally embedded gapping**
Pero el chico la ama y dicen **que** ella ___ a él.
But the boy her loves and they.say that she DOM him
'But the boy loves her and they say that she loves him.' (corpus example from Garcia-Marchena 2018: 179)

In light of the evidence from Farsi, Hindi, Shingazidja and Hungarian locally embedded gapping (and stripping), as well as from both types of embedded gapping in Georgian, Romanian and Spanish, the current version of the Typological Conjecture must be rejected. In all of these languages, complementizers are possible in embedded gapping. Furthermore, those languages for which there is data in both configurations seem to show a pattern: if a language allows complementizers in locally embedded gapping, it also allows complementizers in globally embedded gapping, and vice-versa. This is the case in the Georgian, Romanian and Spanish examples above. Another language that belongs in this group is

Russian, which likewise allows the complementizer in both types of embedded gapping. However, since the patterns in Russian are somewhat subtle and there is disagreement on basic facts in the literature, I will postpone a detailed discussion of Russian until Section 2.4. Consequently, in the (admittedly small) sample of languages discussed so far, there is no language that allows complementizers in one configuration while disallowing them in the other. As a consequence, one might assume a conditional universal underlying the availability of complementizers in embedded gapping rather than a categorical ban on complementizers. This is formulated in Version II of the Typological Conjecture in (89).

(89) **Typological Conjecture on embedded gapping (Version II, to be revised)**
The elliptical conjunct in local embedding can host a complementizer if the elliptical conjunct in global embedding can host a complementizer (and vice versa).

This version of the conjecture requires another small adjustment to be workable. As shown above, Farsi, Hindi and Shingazidja allow complementizers in locally embedded gapping (and Hungarian does so in locally embedded stripping). In its current version, the conjecture predicts that these languages also allow complementizers in globally embedded gapping (stripping for Hungarian). Unfortunately, I have no data on this issue for Hindi and Shingazidja. For Hungarian, Repp (2009: 226) reports that the complementizer *hogy* 'that' is ungrammatical in globally embedded gapping, see (90). Since gapping and stripping show similar embedding behaviour in other languages (i.e. German and English, see Section 2.1), it is possible that *hogy* is also ungrammatical in globally embedded stripping. If this is the case, Hungarian constitutes a counterexample to the Version II of the Typological Conjecture and thus falsifies it. More research on the complementizer in Hungarian globally embedded stripping is needed to clear up this situation.[21]

(90) **Hungarian, globally embedded gapping**
Tudod, hogy János meglátogatta-e a lányokat és (***hogy**) Péter ___ a fiúkat.
Know.2SG[sic!] that Janos visited-INTERROG the girls and that Peter the boys
'I know whether John visited the girls and Peter the boys.' (Repp 2009: 226, I added the judgment on the complementizer, which follows from Repp's discussion of this example.)

In any case, no further speculation as to Hungarian data is required to falsify Version II of the Typological Conjecture: in Farsi, complementizers are not possible in globally embedded gapping, see (91). This means that Farsi allows complementizers in locally embedded gapping, but disallows them in globally embedded gapping, which means that Version II of the Typological Conjecture cannot be accurate.

(91) **Farsi, globally embedded gapping**
Rādmehr midune ke Rātā emruz bāyad bere sare kār va (***ke**) Mehrān fardā ___.
Rodmehr knows that Rata today should go head work and that Mehran tomorrow
'Rodmehr knows that Rata has to go to work today and that Mehran has to go to work tomorrow.' (Farudi 2013: 78f.)

However, this ban on complementizers in Farsi global embedding is not particular to gapping. In the context of coordinate embedded clauses, Farsi never allows a complementizer in non-initial conjuncts,

[21] Notice also that the Hungarian example in (90) contains an embedded interrogative clause, and not an embedded declarative clause like in all the other examples I have hitherto discussed. It is unclear whether the ungrammaticality of *hogy* is particular to interrogative embedding or obtains in declarative clauses as well. This further adds to the unclear empirical situation regarding Hungarian.

whether they be elliptical or not. This is illustrated in (92), where the complementizer *ke* 'that' cannot surface in the second, non-elliptical conjunct.

(92) **Farsi, globally embedded full clause**
midunam ke massie bāhushe va (*ke) maryam xoshgele
I.know that Massie is.intelligent and that Maryam is.pretty
'I know that Massie is intelligent and that Maryam is pretty.' (Farudi 2013: 80)

Since my aim is to formulate a conjecture about complementizers specifically in gapping environments, effects that are independent of ellipsis should be set aside. (See also the discussion of verb effects on the omission of complementizers in English and German local embedding in Section 2.1.) I integrate this aspect of ellipsis-specificity in the final version of the Typological Conjecture in (93).

(93) **Typological Conjecture on embedded gapping (final version)**
Any constraint on a complementizer specific to ellipsis that operates in local embedding also operates in global embedding (and vice versa).

The final version of the Typological Conjecture in (93) covers the cross-linguistic data presented so far. In English, German, Dutch and French, the complementizer is ruled out in both embedded gapping configurations. In Romanian, Georgian and Spanish, the complementizer is acceptable in both configurations. In Farsi, locally embedded gapping allows the complementizer, but globally embedded gapping does not. This is due to independent constraints, which are unrelated to ellipsis. Due to the lack of data on global embedding, it is unclear how Hindi, Shingazidja and Hungarian relate to the conjecture.

One final piece of evidence from the theoretic literature comes from Fernández-Sánchez (2021) on Polish. This language seems to show a categorical difference between locally and globally embedded gapping: in locally embedded gapping, the complementizer *że* 'that' is possible, see (94), whereas in globally embedded gapping, *że* is ungrammatical, see (95).

(94) **Polish, locally embedded gapping**
Łukasz pojechał do Tajlandii, a zgaduję, **że** jego brat ___ do Berlina.
Łukasz travelled to Thailand and guess that his brother to Berlin
'Łukasz travelled to Thailand and I guess that his brother travelled to Berlin.' (Fernández-Sánchez 2021: 115)

(95) **Polish, globally embedded gapping**
Przypuszczam, że Łukasz kupił stary samochód, a (***że**) Maciek ___ rower.
suppose that Łukasz bought old car and that Maciek bicycle
'I guess that Łukasz bought an old car and that Maciek bought a bike.' (Fernández-Sánchez 2021: 122)

If Fernández-Sánchez's (2021) judgments are accurate, they falsify the final version of the Typological Conjecture. However, according to my informant, these judgments only provide a partial picture on embedded gapping in Polish.[22] While the judgment on globally embedded gapping with *że* is not outright incorrect, there is a way to repair this structure. The sentence in (95) is much better if the coordinating conjunction *i* 'and' is substituted for *a* 'and', and there is a pause between conjuncts. My informant also

[22] I thank Łukasz Jędrzejowski for discussing the Polish examples in this section with me.

reports a similar ameliorating effect of coordinating *i* instead of *a* in the local embedding example (94), which with *a* does not sound fully natural to him. What is more, in globally embedded stripping with the polarity particle *też* 'too' and coordinating *i*, *że* is fully acceptable, like in (96).

(96) **Polish, globally embedded stripping**

Przypuszczam, że Łukasz kupił stary samochód, i **że** Maciek ___ też.
suppose that Łukasz bought old car and that Maciek too
'I guess that Łukasz bought an old car and that Maciek bought an old car, too.'

The situation of embedded gapping in Polish seems to closely resemble that in Russian. In Russian, a language closely related to Polish, the complementizer is degraded, but not outright unacceptable in globally embedded gapping (Bonke forthcoming). Fernández-Sánchez's (2021) judgments might reflect a similar degradation in Polish, which was possibly misanalysed as outright ungrammaticality. Furthermore, in Russian there is a sensitivity of the complementizer to the coordinating conjunction like there seems to be in Polish. I go into detail on Russian in Section 2.4. In that section I also discuss the issue of ellipsis-specificity, for which I have no data on Polish, and which I assume to be similar in Russian. Furthermore, I discuss the difference between the two variants of 'and' (*i* vs. *a*), which Russian and Polish both have. All in all, then, Fernández-Sánchez's (2021) judgments on Polish should be treated with caution. This need for a cautious approach is augmented by the fact that Fernández-Sánchez's (2021) judgments on Spanish are incorrect (see above in this section).

In sum, the final version of the Typological Conjecture captures the patterns in the languages discussed in this section. Nevertheless there are some problems with it. First, the empirical situation the conjecture is built on is not great: the available cross-linguistic data on complementizers in embedded gapping is sparse in general, for some languages data is only available on local embedding, and most of the data is based on introspective judgments, some of which are dubious. Furthermore, the conjecture relies on the somewhat vague notion of "complementizer", which might subsume elements under the same term that show great variation between languages despite sharing the same general function as the head of a complement clause. What is more, the conjecture cannot cover all languages, since not all languages even have complementizers to begin with.[23] For these reasons, I do not intend to claim that the "Typological Conjecture" is a universal constraint that holds in every single language.

Instead, I would like to put forward the Typological Conjecture as a new angle on embedded gapping which the literature usually does not take. The final version of the conjecture in (93), reproduced here in (97) for convenience, does not only deal with the outright grammaticality or ungrammaticality of complementizers, but speaks more broadly of general "constraints". Outright ungrammaticality is of course also a constraint, but this more general formulation also allows for the inclusion of soft constraints that do not result in outright ungrammaticality. This inclusion of soft constraints opens up a new avenue of more fine grained investigation into globally and locally embedded gapping.

(97) **Typological Conjecture on embedded gapping (final version)**

Any constraint on a complementizer specific to ellipsis that operates in local embedding also operates in global embedding (and vice versa).

[23] For a discussion of embedded gapping in a language where clausal embedding proceeds via nominalization and without complementizers, see İnce 2009 on Turkish.

In the following section, I outline one way in which this version of the Typological Conjecture may lead to fruitful results in ellipsis research. I will illustrate what a fine-grained investigation of the Typological Conjecture may look like, by focussing on a language that displays soft constraints on the complementizer in locally and globally embedded gapping: Russian.

2.4 Discourse relations in embedded gapping: Russian

In the previous section, I made the conjecture that restrictions on complementizers are identical in locally and globally embedded gapping across languages. The approach I took to make this conjecture was coarse grained: I looked at whether the presence of complementizers is generally grammatical or ungrammatical. However, I formulated the conjecture so that it potentially also covers soft constraints, albeit without providing any evidence. In the present section, I take a more fine-grained approach, by outlining a case study on Russian. In this language, we find the desired evidence for soft constraints on the complementizer *čto* 'that' in globally and locally embedded gapping. The discussion of globally embedded gapping is from Bonke (forthcoming), which contains robust quantitative data. For locally embedded gapping, there is no such robust quantitative data (yet). Instead, I will discuss two pieces of tentative evidence: one is an interpretation of the available introspective acceptability judgment data from the literature, the other is the results of a small-scale acceptability experiment.[24]

The only publication I am aware of that briefly touches upon *čto* in globally embedded gapping is Repp (2009), which reports the introspective judgment that the complementizer is ungrammatical, see (98). This judgment does not capture the full behaviour of *čto*: quantitative data reveals that the acceptability of *čto* is more nuanced than Repp's (2009) judgment suggests and subject to subtle grammatical constraints. Specifically, the results of four acceptability experiments that are reported in Bonke (forthcoming) reveal that globally embedded gapping with *čto* is degraded in comparison to the variant without *čto*, which in broad terms is congruent with Repp's (2009) judgment. However, acceptability ratings are not so low as to indicate outright unacceptability. The results also indicate that further constraints on *čto* relate to the way the two conjuncts are connected.

(98) Ja dumaju, čto Petr ezdit v Indiju a (***čto**) Pavel ___ v Švejcariju.
I think that Peter is.going to India and that Pavel to Switzerland
'I think that Peter is going to India and that Pavel is going to Switzerland.' (Repp 2009: 226)

Russian has two coordinating conjunctions, *a* like in (98) and *i*, which both roughly correspond to English *and*. I will not go into the details of the semantics and usage conditions of *i* vs. *a*, see Mendoza (1996), Jasinskaja & Zeevat (2008, 2009), amongst others. I make the simplifying generalization that *a* indicates a contrastive relation between the conjuncts (cp. Tauscher & Kirschbaum 1970), while *i* indicates a parallel (Mendoza 1996), non-contrastive relation. I use "contrastiveness" as an auxiliary notion that for the sentences discussed here comes down to whether the VPs in the coordinate clauses are identical

[24]This section has benefited from the discussion with native Russian speakers. In this regard, I would like to thank Solomeya Bagautdinova, Ljudmila Geist, Katja Jasinskaja, Jenia Khristoforova, Anastasiia Kurbaeva, and Alina Yehorova. I also thank Boban Arsenijević for his comments and discussion.

(= non-contrastive, coordinator *i*) or non-identical (= contrastive, coordinator *a*).[25] Example (98) above is contrastive in this sense: the VPs are non-identical because of the different PPs *to India* vs. *to Switzerland*. An example of what a non-contrastive coordination would look like is the stripping example in (99). Here, the VP *is going to India* in the first conjunct is identical to the deleted VP in the second conjunct, which licenses coordinating *i*.

(99) Ja dumaju, čto Petr ezdit v Indiju i (**čto**) Pavel tože ~~ezdit~~ ~~v~~ ~~Indiju~~.
I think that Peter is.going to India and that Pavel too is.going to India
'I think that Peter is going to India and that Pavel is going to India, too.'

In contrastive coordination with *a* in globally embedded gapping, *čto* leads to the aforementioned degradation in ratings (Bonke forthcoming). In non-contrastive coordination with *i*, *čto* leads to no such degradation. However, with *i*, ratings indicate a general degradation, independently of *čto*. For a detailed description and a sketched analysis of these patterns, see Bonke (forthcoming). What is relevant for the discussion at hand is that these patterns in globally embedded gapping make predictions on locally embedded gapping. Specifically, the Typological Conjecture predicts that we should find the same sensitivity to the discourse relation between conjuncts (as indicated by the coordinating conjunction) in locally embedded gapping that we find in globally embedded gapping.

As in the gapping literature generally, local embedding has received more attention than global embedding in the literature on Russian. The three sentences in (100) to (102) constitute an exhaustive list of examples. Notice the disagreement on acceptability: (100) and (101) are judged to be unacceptable, while Erschler (2018: 64, fn. 29) considers his example (102) acceptable in colloquial speech. Erschler (2018: 64, fn. 29) acknowledges this disagreement in acceptability judgments but does not further discuss it.

(100) * Maša budet čitat' knigu, a Lena dumala, **čto** Ivan ___ gazetu.
Masha will read book and Lena thought that Ivan newspaper
'Masha will be reading a book and Lena thought that Ivan will be reading a newspaper.'
(Grebenyova 2012: 69)

(101) * Vasja poedet v Moskvu, i ja dumaju, **čto** Petja ___ v Peterburg.
Vasja will.go to Moscow and I think that Pete to St. Petersburg
'Vasja will go to Moscow and I think that Pete will go to St. Petersburg.' (Kazenin 2010: 88, Bailyn & Bondarenko 2018: 1010)

(102) Vasja p'jot samogon i mne kažetsja **što**[sic] Oleg ___ vodku.
Vasja drinks moonshine and to.me seems that Oleg vodka
'Vasja drinks moonshine and it seems to me that Oleg drinks vodka.' (Erschler 2018: 5, 65)

The basic empirical facts on locally embedded gapping in Russian thus do not seem straightforward and further research is required for potential factors influencing its acceptability. I will discuss one such potential factor here, which based on (100) to (102) may be the coordinating conjunction. In this discussion,

[25] Another way to describe the same facts would be in terms of single vs. multiple contrast (e.g. Bassi & Bondarenko 2021): When the coordinate clauses contain exactly one pair of contrasting elements (e.g. the subjects in 99) the coordinator is *i*. If there is more than one pair of contrasting elements (e.g. the subjects and the PPs in 98), the coordinator is *a*. To me, the term "contrastive" does not seem quite right to describe the relation between the non-identical constituents (= the embedded subjects) in coordinate clauses like (99), since the situations their referents are situated in (= the embedded VPs) are identical. Therefore, I will not adopt this terminology.

I will set aside the potential impact of the presence or absence of *čto*, since the existing literature provides no insights on this issue. All Russian examples of local embedding presented in this section so far have featured *čto*, which is why I will continue the discussion on the basis of examples that have *čto*. Generally speaking, a complement clause headed by *čto* seems to be the syntactically unmarked variant compared to *čto*-less clauses. I return to the issue of complementizer optionality in Russian in Section 2.4.2.

The coordinating conjunction varies between the examples in (100) to (102). In (100) the coordinator is *a*, in (101) and (102) it is *i*. This difference should be interpreted with caution, given that two of the examples are judged to be unacceptable. But it is nevertheless striking that (101) and the acceptable (102) contain *i*, even though the VPs in the two conjuncts are contrastive in the sense outlined above and should therefore require *a*. Without the embedding clause, the sentences in (100) to (102) are only fully acceptable with *a*; compare for instance the non-embedded (103) to its embedded equivalent (101) above. However, coordinator selection seems to be a soft constraint, since the version of (103) with *i* is not outright unacceptable.[26]

(103) Vasja poedet v Moskvu, {a / ??i } Petja v Peterburg.
Vasja will.go to Moscow and and Pete to St. Petersburg
'Vasja will go to Moscow and Pete to St. Petersburg.'

The divergent judgments in (100) to (102) might nevertheless be explained by assuming that like *čto* in globally embedded gapping, locally embedded gapping is much better with coordinating *i*. This assumption straightforwardly explains the unacceptability of (100), because the coordinator is *a*. It equally straightforwardly explains the acceptability of (102), because it has the correct coordinator *i*. The problematic example is (101) because it is judged unacceptable even though it has *i*. In the contrast between (101) and (102), the soft coordinator constraint might be at play: for Kazenin (2010) and Bailyn & Bondarenko (2018), the coordinator *i* in a contrastive context where *a* is generally preferred might cause a degradation in acceptability that is enough to indicate unacceptability for their example (101). For Erschler (2018), there may also be a degradation, which however does not constitute outright unacceptability. I take Erschler's (2018) remark that locally embedded gapping is limited to informal registers as tentative evidence of this degradation.

To provide more data on locally embedded gapping and its interaction with the coordinator, and to double-check the introspective judgments in (100) to (102), I conducted an acceptability judgment experiment. The primary function of this experiment was to serve as a pre-test for the lexical materials in the experiments from Bonke (forthcoming), and as a test on the coordinators *i* and *a* in non-embedded contexts. I report on the results of this pre-test experiment in Appendix B. What is crucial for the present discussion is the fact that this pre-test experiment contained filler items with locally embedded gapping (and stripping). I report the results of these filler items in the following section. These results are not reported in Bonke (forthcoming).

[26] See the Appendix B for quantitative evidence on the difference between *a* and *i* in gapping that is indicated by the judgments in (103).

2.4.1 An empirical exploration of the coordinators *i* and *a* in locally embedded gapping and stripping in Russian

The experimental materials consisted of four test items, which were based on the three examples from the literature in (100) to (102) above, plus one constructed item. The experiment had a 2 × 2 design with the factors coordinator (*i* vs. *a*) and ellipsis type. The two ellipsis types were gapping, and stripping with the polarity particle *tože* 'too'. The four items are in (104) in all four conditions. In the items taken from the literature (104a) to (104c), I changed some of the proper names to avoid repetition and adjusted the punctuation and spelling where necessary so as to conform to the standard. Due to an oversight, the constructed item in (104d) contained a typo: the embedding verb in should be *predpolagaet*. This did not seem to affect ratings, which is why I did not exclude it from the analysis (see Table 2.2 and Figure 2.2 further below).

(104) **Experimental items in all four conditions** (without underlining and boldface in the actual questionnaire)

a. Maša budet čitat' knigu, {i / a } Lena dumala, **čto** Ivan – {gazetu / tože}.
Masha will read book and and Lena thought that Ivan newspaper too
'Masha will be reading a book and Lena thought that Ivan will be reading a newspaper/will be reading a book, too.' (after Grebenyova 2012: 69)

b. Vova poedet v Moskvu, {i / a } ja dumaju, **čto** Petja – {v Peterburg / tože}.
Vova will.go to Moscow and and I think that Pete to St. Petersburg too
'Vova will go to Moscow and I think that Pete will go to St. Petersburg/will go to Moscow, too.' (after Kazenin 2010: 88 and Bailyn & Bondarenko 2018: 1010)

c. Vasja p'jot samogon {i / a } mne kažetsja, **čto** Oleg – {vodku / tože}.
Vasja drinks moonshine and and to.me seems that Oleg vodka too
'Vasja drinks moonshine and it seems to me that Oleg drinks vodka/drinks moonshine, too.' (after Erschler 2018: 5, 65)

d. Paša kupil rozy, {i / a } Il'ja predpologaet,[sic] **čto** Senja – {fialki / tože}.
Pasha bought roses and and Ilja assumes that Senja violets too
'Pasha bought roses, and Ilja assumes that Senja bought violets/bought roses, too.' (based on a non-embedded gapping example from Kalinin 2020: 4)

If my conclusions from the theoretic discussion above are correct, gapping should not be fully acceptable with either coordinator: with *a*, embedded gapping is not licensed, and *i* is the wrong coordinator for contrasting conjuncts. Both of these constraints are (by assumption) soft constraints, so no outright unacceptability is predicted either. The inclusion of stripping with *tože* 'too' is crucial in the design because its conjuncts are non-contrastive and therefore the preferred coordinator is *i*. Under the assumption that gapping and stripping behave otherwise identically (see Section 2.1), locally embedded stripping with *i* is predicted to be acceptable because *i* licenses embedded gapping. Stripping with *a* is predicted to be degraded because *a* is not the appropriate coordinator.

The four items were distributed across four lists, so that each participant saw all four items in a unique condition each (= Latin square). In addition, each list contained 24 items from another experiment which tested coordinator selection in non-embedded gapping and *tože*-stripping, plus four unacceptable controls items. 36 participants completed the online questionnaire. They judged the naturalness of the items

on a continuous scale. The extremes of the scale were labelled *očen' neestestvenno* 'very unnatural' and *očen' estestvenno* 'very natural'. The resulting slider positions indicating the participants' naturalness judgments for the individual items were coded as integers from 0 to 100. This procedure was identical to that in Bonke (forthcoming)

Due to the small scale of the experiment, the results should be interpreted with caution. The results are are summarized in Table 2.1 and Figure 2.1.[27] One of the control items received a mean rating of 60.5, which I deemed too high to be indicative for unacceptability. This item did not enter the analysis. The data from all other items (4 experimental, 3 controls) and all participants entered the analysis.

Table 2.1: Experimental results for locally embedded gapping in Russian (*N* = number of observations, *SD* = standard deviation, *SE* = standard error, *CI* = 95% confidence interval).

Ellipsis Type	Coordinator	N	Mean rating	SD	SE	CI
Gapping	*a*	36	45.11	28.05	4.68	9.49
Gapping	*i*	36	40.42	30.81	5.14	10.42
Stripping	*a*	36	37.08	24.54	4.09	8.30
Stripping	*i*	36	80.11	29.69	4.95	10.05
Unacceptable controls		108	19.64	22.95	2.21	4.38

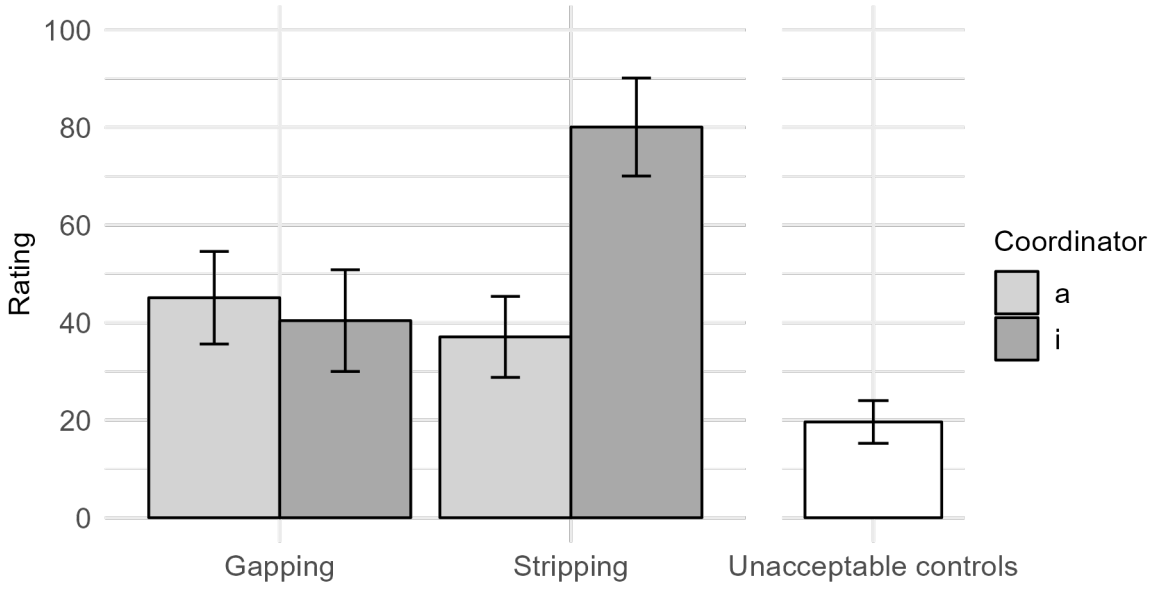


Figure 2.1: Results for Russian locally embedded gapping and unacceptable controls. Bars indicate mean values; error bars indicate 95% confidence intervals.

Only stripping with *i* receives comparably high ratings. The other three conditions receive degraded ratings. The differences between these three conditions are only slight, which does not suggest that there is a categorical difference between them in acceptability. There is, however, a noticeable difference between the three degraded conditions and the unacceptable controls. I interpret this difference to mean that the degradation in the conditions is not so large as to indicate outright unacceptability. The ratings for the experimental conditions are thus as the Typological Conjecture predicted.

[27] I made the plot in `R` with the `ggplot2`-package (Wickham 2016). The values in Table 2.1 were calculated with the function `summarySEwithin()` from the `Rmisc`-package (Hope 2022). This function removes inter-subject variability.

I think this descriptive analysis of the experimental data suffices to paint an informative picture of my data on locally embedded gapping in Russian. For sake of completeness, I also report the results of an inferential analysis. I used R (R Core Team 2023) and the afex-package (Singmann et al. 2022) to fit a linear mixed effects model to the data. The fixed effects were the ellipsis type (gapping vs. stripping), the coordinator (*i* vs. *a*) and the interaction. The model included random intercepts for items and participants. (More complex random structures did not converge, probably due to the low number of observations and/or items.) Predictors were sum-to-zero contrast coded to facilitate the interpretation of interaction effects. The model indicated significant effects of the ellipsis type ($b = 7.9, SE = 2.3, t = 3.5, p < 0.001$) and the coordinator ($b = 9.6, SE = 2.3, t = 4.2, p < 0.001$). The interaction effect was significant as well ($b = 11.9, SE = 2.3, t = 5.2, p < 0.001$). To resolve the interaction, I split the data into a gapping set and a stripping set and fit a linear model each. These models each had the coordinator as the sole fixed factor and random intercepts for participants and items. The stripping model indicated a significant effect of the coordinator ($b = 21.5, SE = 3.0, t = 7.3, p < 0.001$). In the gapping model, the effect of the coordinator was not significant ($b = -2.3, SE = 3.6, t = -0.7, p = 0.5$). This suggests that locally embedded *tože*-stripping is indeed more acceptable with *i* than with *a*, while for gapping, the difference in coordinator is not relevant.

Finally, Table 2.2 and Figure 2.2 show the ratings for the individual items (pooled for all conditions). While there are some considerable differences between the mean ratings of the four items, the confidence intervals overlap. I take these overlapping confidence intervals to indicate that the differences between the items are no cause for concern. These findings justify the non-exclusion of the typo-item (104d) (which is not even the worst rated item overall) from the analysis.

Table 2.2: Mean ratings for the individual items (all conditions).

Item	N	Mean rating	SD	SE	CI
104a (*Masha …*)	36	40.06	27.56	4.60	9.33
104b (*Vova …*)	36	52.42	34.10	5.68	11.54
104c (*Vasja …*)	36	62.39	38.15	6.36	12.91
104d (*Pasha …*)	36	47.86	34.30	5.72	11.60

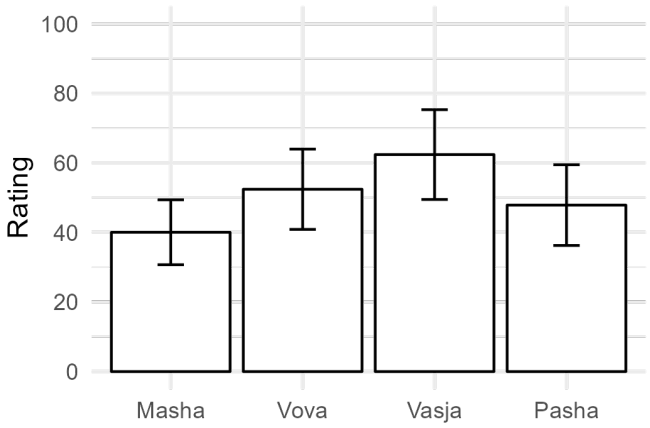


Figure 2.2: Ratings for individual items (by matrix subject). Mean ratings with 95% confidence interval.

2.4.2 Analysis and conclusions on Russian

As indicated above, the data on Russian locally and globally embedded gapping presented in this section may at first glance be interpreted in favour of the Typological Conjecture: in both configurations, the coordinating conjunction and the discourse relation between the conjuncts that comes with it constrain acceptability. If the conjuncts are connected with the contrastive coordinator *a*, there is a degradation for *čto* in globally embedded gapping (Bonke forthcoming), and for locally embedded gapping as a whole. If the conjuncts are coordinated with non-contrastive *i*, there is no such degradation.

However, there are some confounding factors that limit the way in which the Russian data can be adduced as evidence in favour of the Typological Conjecture. First, the evidence for the coordinator effect in locally embedded gapping comes from introspective judgments in the available literature and a small-scale experiment. It is therefore tentative; more study is needed.

Furthermore, in Bonke (forthcoming), I find that the effect of the coordinator in global embedding is not particular to elliptical clauses: *čto* is sensitive to the distinction of *i* vs. *a* in non-elliptical coordinate clauses as well.[28] Thus, the negative effect of *a* on the acceptability of *čto* must thus be a general constraint in Russian (see also Bassi & Bondarenko 2021). Recall that I introduced an aspect of ellipsis-specificity into the "Typological Conjecture" above to account for the Farsi data (see Section 2.3 above). This aspect now rules out the Russian data as evidence for the Typological Conjecture, since the observed effects are not specific to ellipsis.

Finally, as highlighted at the beginning of this section, the discussion on locally embedded gapping center around the structure as such and ignores the role of the complementizer within that structure. All examples of locally embedded gapping, whether they came from the literature or were tested in the present experiment, contained *čto*. I have (implicitly) made the assumption that like in French (see Section 2.3), the complementizer is the element that triggers the degradation, which then "projects" to the sentence as a whole. Unlike in French, though, the general prerequisite for locally embedded clauses without *čto* is met, since Russian allows complementizer omission. However, complementizer omission is restricted in multiple ways, which may provide difficulties for an experimental examination of complementizerless locally embedded gapping: it is restricted to colloquial registers (King 1993, Hansen et al. 2016) and subject to (possibly interconnected) pragmatic (Hansen et al. 2016) and verb type constraints (Knyazev 2022). Possible future research aiming at pinpointing the specific role of *čto* in locally embedded gapping should proceed under careful consideration of these independent constraints on *čto*-omission.

In sum, more research is needed to shed more light on the interaction of the complementizer, coordinator choice and ellipsis in Russian. More importantly, even though the evidence presented in this Section cannot be adduced as evidence for the Typological Conjecture, it certainly does not falsify it either. If anything, in this Section I have hoped to show how the Typological Conjecture may provide a new perspective on the study of the interaction of gapping and embedding and, which leads to interesting results, even though it does not provide evidence for or against the conjecture itself.

[28] I have no data on this for local embedding.

2.5 Factivity: Spanish

There are two reasons that Spanish provides an ideal testing ground for the Typological Conjecture: first, the complementizer *que* 'that' is generally possible in globally embedded gapping (*pace* Fernández-Sánchez 2021), which is shown in the corpus examples in (105) to (107). See also example (87) in Section 2.3 above for another corpus example with the same pattern. The fact that *que* in embedded gapping surfaces in newspapers and legal texts suggests that it is not limited to colloquial registers. Furthermore, *que* does not seem to be obligatory in globally embedded gapping, see example (108), in which the elliptical clause contains no *que*.[29]

(105) Mercado Libre [...] detectó que el 36,8% de los compradores tienen entre 25 y 34 años
market free found that the 36,8% of the shoppers have between 25 and 34 years
y **que** el 13,4%, ___ entre 18 y 24 años.
and that the 13,4% between 18 and 24 years
'Mercado Libre found that 36.8% of shoppers are between 25 and 34 years old and that 13.4% are between 18 and 24 years old.' (from an Argentinian online newspaper)[30]

(106) [E]l sistema [...] ocasiona que las cervazas más baratas paguen más impuestos y **que** las
the system causes that the beers most cheap pay more taxes and that the
mas caras ___ menos.
most expensive less
'The system results in the cheapest beers paying more tax and the most expensive ones paying less tax.' (from a Peruvian online newspaper)[31]

(107) Advierte que las faltas gravísimas presuponen el dolo y **que** las restantes, ___
notes that the misdemeanours serious-SUP presuppose the malice and that the remaining
el elemento culposo.
the element culpable
'[The Administrative Department of the Civil Service] notes that the most serious misdemeanours presuppose malice and that the remaining misdemeanours presuppose the element of guilt.' (from a Colombian legal text)[32]

(108) [M]e informó que el alto era mi sistema nervioso, y el bajo ___ mi sangre en
me informed that the high was my system nervous and the low my blood in
circulación.
circulation
'He informed me that the high [sound] was my nervous system and the low one was my circulating blood.' (from a Spanish online newspaper)[33]

[29] In addition to gapping, the examples in (105) to (108) all involve ellipsis within the subject DPs of the elliptical conjuncts. In (105), the full subject would be *13.4% [of the shoppers]*, in (106) *the most expensive [beers]*, in (107) *the remaining [misdemeanors]* and in (108) *the low [sound]* (here, the subject DP in the first clause is elliptical, too). This is an interesting pattern, but seems to be coincidental. There is no obvious reason why DP-internal ellipsis and gapping should be connected. The corpus example (87) in Section 2.3 involves no DP-internal ellipsis. Notice furthermore that all corpus examples with *que* (including 87 in Section 2.3) stem from American sources, while the one without *que* in (108) is from Spain. I assume that this is coincidental as well.

[30] www.jornadaonline.com//economia/176773-hot-sale-furor-por-pasajes-a-chile, accessed 12 November 2019. Mercado Libre is an Argentinian online retailer.

[31] www.gestion.pe/economia/sigue-controversia-calculo-isc-aplicado-cervezas-43073-noticia, accessed 12 November 2019.

[32] www.corteconstitucional.gov.co/relatoria/2002/C-181-02.htm, accessed 21 March 2024.

[33] www.elcorreoweb.es/maspasion/guia-cofrade-silencio-383-FD3894986, accessed 20 May 2020.

This situation provides the necessary prerequisites for testing the Typological Conjecture: the complementizer is not ruled out altogether (like in English, German and Farsi), which means that the examination of *que* for further constraints is possible in the first place. The complementizer is not obligatory either, which means that any pattern of *que* in globally embedded gapping can be checked against its unmarked, i.e. *que*-less, counterpart. If we compare the versions with vs. without *que*, we will be able to tell whether the observed constraints are indeed particular to *que* or a general property of elliptical or coordinated clauses.

The second reason to look at Spanish is that locally embedded gapping is not only possible (see Section 2.3 for examples), but also displays the type of ellipsis-specific constraints that I am interested in. In Spanish, these constraints pertain to the type of the embedding verb; I will discuss them in detail in Section 2.5.1. What makes Spanish an even better subject still for investigation is that the empirical situation on locally embedded gapping is comparatively good: the recent-ish quantitative study by Bîlbîie & de la Fuente (2019) provides detailed and robust quantitative data on verb type constraints. The Typological Conjecture predicts that the same constraints that Bîlbîie & de la Fuente (2019) find for locally embedded gapping work on *que* in globally embedded gapping. I will test this prediction in Chapter 3.

There is, however, a confound. The comparison I draw is between *que* in globally embedded gapping and locally embedded gapping as such. Within locally embedded gapping, it is difficult to disentangle the individual role that *que* might play in the verb type constraints from other potential influencing factors (whatever these may be). This is because Spanish does not easily allow complementizer drop in "regular" (i.e. non-coordinated, non-elliptical) complement clauses (cp. the discussions of French in Section 2.3 and Russian in Section 2.4.2). This means that we cannot easily compare locally embedded gapping with *que* to its counterpart without *que* (like we did for German, English and Dutch in Sections 2.1 and 2.3), as most complement clauses are ungrammatical without *que*, independently of ellipsis.

Although the omission of *que* from a complement clause is not outright impossible as sometimes claimed (Bîlbîie et al. 2023: 421), the it is subject to various types of constraints, the exact dimensions and interconnections of which do not seem to be totally clear. The structure of the embedded clause seems to play a role, where omitted *que* is possible in subjunctive clauses that have no explicit (preverbal) subject. This is the case in example (109), where there is no explicit second person pronoun accompanying the subjunctive finite verb *sigas* 'keep'. In this sentence, *que* may be omitted.

(109) Me alegra (**que**) sigas mejorando de salud.
me pleases that keep-SUBJ-2SG improving of health
'I am glad that your health is still improving.' (Torrego 1983: 562, my gloss)

There is no consensus as to the precise extent to which the subjunctive mood on the embedded verb plays a role in *que*-omission. In some publications, only a sub-set of subjunctive-taking embedding predicates are claimed to allow *que*-omission (verbs of fear, wish and volition, Delbecque & Lamiroy 2000), in others any subjunctive-taking predicate allows *que*-omission (Torrego 1983, Kempchinsky 1998), and in still others, any form of the embedded verb with an irrealis meaning (i.e. the subjunctive, but also the future tense or conditional) licenses *que*-omission (Brovetto 2002). In any case, omitted *que* also seems to be limited to formal written registers (Torrego 1983, Mazzola 2022) and is subject to dialectal variation (Mazzola 2022).

An in-depth discussion of *que*-omission is beyond the scope of this dissertation. For a recent detailed study, see Mazzola (2022). What is relevant at this point is that *que*-omission is so heavily constrained that an examination of locally embedded gapping without *que* is not feasible. As a consequence, all instances of locally embedded gapping discussed in this dissertation will feature *que*. This procedure raises the empirical problem of whether it is actually *que* that triggers the verb type effects in locally embedded gapping that we will see in the following section. To overcome this problem, I will simply assume that this is the case, i.e. that constraints on locally embedded gapping with *que* are constraints on *que* itself (much like I did for French in Section 2.3). As we will see in Bîlbîie & de la Fuente's (2019) analysis of their data (Section 2.5.3), they do not share this assumption. What is crucial for me, though, is that this procedure motivates the comparison of locally embedded gapping to *que* in globally embedded gapping under the Typological Conjecture.

In the following section, I summarize the experiments and discussion in Bîlbîie & de la Fuente's (2019) study on restrictions on locally embedded gapping. A central aspect of this study is factivity, which will play a central role in this dissertation as well, see Chapter 3.1.1 where I discuss factivity in depth. For now, I uncritically follow Bîlbîie & de la Fuente's (2019) assumptions on the matter. Parts of the following sections are from Bonke & Repp (2022).

2.5.1 Locally embedded gapping and factivity in Spanish

The reason to assume that factivity plays a role in the acceptability of locally embedded gapping comes from the theoretic literature. Fernández-Sánchez (2021) reports that non-factive embedding predicates such as *creo* 'think' or *imagino* 'imagine' license gapping in their complement clause, see (110a). Factive predicates like *lamento* 'regret' or *odio* 'hate' on the other hand do not allow gapping in their complement clause, see (110b). Similarly, de Cuba & MacDonald (2013) report that fragments can be embedded under non-factive embedding predicates, see (111a), but not under factive predicates, see (111b).[34] There is some quantitative evidence for this verb type restriction: Garcia-Marchena's (2018) corpus study finds only epistemic, non-factive embedding predicates for both embedded fragments (Garcia-Marchena 2018: 174) and locally embedded gapping (Garcia-Marchena 2018: 179). Recall that the embedding facts in English suggest that locally embedded gapping and embedded fragments may receive a unified treatment, see Section 2.1.

(110) a. Alfonso robó las esmeraldas y {creo / imagino / ...} **que** Mugsy ___ las perlas.
Alfonso stole the emeralds and think imagine that Mugsy the pearls
'Alfonso stole the emeralds and I think/imagine that Mugsy stole the pearls'
(Fernández-Sánchez 2021: 119)

b. * Alfonso robó las esmeraldas y {lamento / odio / ...} **que** Mugsy ___ las perlas.
Alfonso stole the emeralds and regret hate that Mugsy the pearls
'Alfonso stole the emeralds and I regret/hate that Mugsy stole the pearls'
(Fernández-Sánchez 2021: 119)

[34] The different diacritic marks in (110b) and (111b) indicate that Fernández-Sánchez (2021) assumes ungrammaticality in locally embedded gapping under factives, while de Cuba & MacDonald (2013) assume that fragments embedded under factives are semantically malformed. I have nothing to contribute on the difference at this point, but note that de Cuba & MacDonald (2013) propose a structural (rather than a semantic) account for their data. I will concentrate on the fact that both examples are ruled out.

(111) ¿Quién robó las joyas?
who stole the jewels
'Who stole the jewels?'

a. {Creo / supongo / me imagino / pienso } **que** tu hijo ___.
believe suppose me imagine think that your son
'I believe/suppose/imagine/think that your son stole the jewels.'

b. # {Lamento / sé / me sorprende / me desagrada } **que** tu hijo ___.
regret know me surprises me displeases that your son
'I regret/know/it surprises me/it displeases me that your son stole the jewels.'

(adapted from de Cuba & MacDonald 2013: 321)

Taking de Cuba & MacDonald's (2013) and Fernández-Sánchez's (2021) judgments as a starting point,[35] Bîlbîie & de la Fuente (2019) investigate locally embedded gapping experimentally. Their main issue with the introspective data is that it conflates two distinct types of embedding predicates under the term 'factive': emotive predicates, such as *lamento* 'regret' or *odio* 'hate', and knowledge/perception predicates, such as *sé* 'know' or *veo* 'see'. This conflation can be seen in de Cuba & MacDonald's (2013) example (111b), which features both types of predicates: the knowledge verb *sé* 'know', and the emotives *lamento* 'regret', *me sorprende* '(it) surprises me' and *me desagrada* '(it) displeases me'.

There are two reasons why these two types of predicates should be kept apart: first, they differ in mood selection. Emotive predicates typically select for a clause which has the embedded finite verb in its subjunctive form, while knowledge and perception predicates select for a clause with the embedded finite verb in the indicative. The difference is illustrated in (112) and (113), where the embedded verb *robar* 'steal' varies between the subjunctive form *robara* and the indicative *robó*, depending on the type of embedding verb (emotive *lamento* 'regret' in 112, knowledge *sé* 'know' in 113). Notice however that mood selection is a preference rather than a steadfast rule, and may alternate in certain environments (see e.g. Lope Blanch 1958, Borrego Nieto et al. 1986, Quer 2011, Faulkner 2021a,b amongst many others). I will return to the issue of mood selection in the discussion of my own experimental materials in Chapter 3.1.2.

(112) Lamento que Mugsy robara las perlas.
regret-IND that Mugsy stole-SUBJ the pearls
'I regret that Mugsy stole the pearls.'

(113) Sé que Mugsy robó las perlas.
know-IND that Mugsy stole-IND the pearls
'I know that Mugsy stole the pearls.'

The difference in mood selection is relevant to locally embedded gapping because it leads to a mismatch: under emotives, the elided verb would appear in the subjunctive (had it not been deleted), and the antecedent verb appears in the indicative, the unmarked mood in main clauses. This mismatch is illustrated in (114), which is the full clause version of Fernández-Sánchez's (2021) example in (110b). The same mismatch does not obtain for knowledge and perception predicates, since they select for indicative, i.e. the same mood as the verb of the antecedent. See (115), the full clause version of Fernández-Sánchez's (2021) example (110a).

[35] Bîlbîie & de la Fuente (2019) cite an earlier, unpublished version of Fernández-Sánchez's (2021) paper, which seems to contain the same judgments.

(114) Mood mismatch (indicative *robó* - subjunctive *robara*)
Alfonso robó las esmeraldas y lamento **que** Mugsy robara las perlas.
Alfonso stole-IND the emeralds and regret that Mugsy stole-SUBJ the pearls
'Alfonso stole the emeralds and I regret that Mugsy stole the pearls.'

(115) Mood match (indicative *robó* - *robó*)
Alfonso robó las esmeraldas y sé **que** Mugsy robó las perlas.
Alfonso stole-IND the emeralds and know that Mugsy stole-IND the pearls
'Alfonso stole the emeralds and I know that Mugsy stole the pearls.'

Thus, the mismatch in mood may be responsible for the ungrammaticality of the gapping and fragments examples in (110b) and (111b) above, rather than semantic or pragmatic properties of the embedding verb like factivity. As is well known, gapping tolerates some mismatches more easily than others. Mismatches concerning agreement features are generally unproblematic: in (116), the verb in the first conjunct agrees with the null singular subject, and the verb in the second conjunct agrees with the plural subject *los mexicanos* 'the Mexicans'. The resultant mismatch in the grammatical forms of the verb *es* 'is' vs. *son* 'are' does not lead to ungrammaticality.

(116) Agreement mismatch (singular *es* vs. plural *son*)
El mundo es su empresa y los mexicanos ~~son~~ sus lacayos.
the world is his company and the Mexicans are his lackeys
'The world is his company and the Mexicans are his lackeys.' (adapted from from a corpus example of globally embedded gapping in Bonke & Repp 2022: 528, cp. 87 above)

At least some non-agreement features of the verb do not tolerate mismatches in gapping. This concerns for instance tense or voice, see (117) and (118), respectively. Given that mood is not an agreement feature, it is not unlikely that it patterns like tense or voice in that it causes ungrammaticality if mismatched.

(117) **Tense mismatch (past *compró* vs. future *comprará*)**
*Felipe compró los libros ayer y Raúl ~~comprará~~ las revistas mañana.
Felipe bought the books yesterday and Raúl will.buy the magazines tomorrow
'Felipe bought the book yesterday and Raúl will buy the magazines tomorrow.' (filler item to the acceptability experiments in Chapter 3.)

(118) **Voice mismatch**
*Some bring roses and lilies ~~were brought~~ by others. (Alves Castro 2016: 178, Lu & Kim 2022: 1)

The second reason to keep the two types of predicates apart is Bîlbîie & de la Fuente's (2019) assumption that the two types of embedding predicates align with two different sub-classes of factives: Emotives are 'true' factives and knowledge/perception verbs are semi-factives. This partition of factive verbs into true factives and semi-factives has been proposed early on in factivity research (e.g. Karttunen 1971, Hooper & Thompson 1973, Hooper 1975), and was originally based on the observation that not all factive predicates project a presupposition in all of the different entailment cancelling environments (Karttunen 1971).[36] In their exposition, Bîlbîie & de la Fuente (2019) largely rely on the suggestive nature of the term "semi-factive", which they seem to interpret as taking up an intermediate position between non-factives and true factives. They offer more specific details on the different properties of the types of embedding predicates

[36] I discuss factivity in more depth in Chapter 3.

in the analysis of their data, to which I will return in Section 2.5.3 below. In the following section, I turn to Bîlbîie & de la Fuente (2019) experimental study.

2.5.2 Experimental evidence on factivity in locally embedded gapping: Bîlbîie & de la Fuente (2019)

With the three-way distinction of embedding predicates (non-factive vs. semi-factive vs. true factive) in mind, Bîlbîie & de la Fuente (2019) conducted two acceptability judgment experiments on locally embedded gapping. Both experiments had the same overall design: they contrasted locally embedded gapping with the three types of predicates against non-embedded gapping on the one hand, and against the corresponding full clauses on the other hand. The statistical analysis involved linear mixed-effects models, in which the two factive sub-types were collapsed into a single factor level 'factive'. This resulted in a 3 × 2 factorial design with the factors embedding (non-embedded vs. non-factive embedding vs. factive embedding) and gapping (gapping vs. full clause). The differences between the two sub-types of factives received a separate statistical analysis, which was descriptive only.

To illustrate the experimental factors, example (119) contains a sample item in the 6 conditions from Experiment 1. There are two things to note in the items for Experiment 1. First, all verbs are in the indicative, including the one in the non-elliptical condition with an emotive factive (119f). In these items, the preferred form on the embedded verb *pedir* 'order' would be the subjunctive *pidiera* rather than the indicative *pidió*. Second, items were introduced by a *wh*-question (*What drinks did the two friends order in the bar?*), which "served as a contextual anchor" (Bîlbîie & de la Fuente 2019: 7).

(119) ¿Qué bebidas pidieron los dos amigos en el bar?
what drinks ordered the two friends in the bar
'What drinks did the two friends order in the bar?'

a. [Gapping, No Embedding]
Pablo pidió una cerveza y Juan ___ un whisky.
Pablo ordered a beer and Juan a whisky
'Pablo ordered a beer and Juan ordered a whisky.'

b. [Gapping, Embedding Non-factive]
Pablo pidió una cerveza y sospecho **que** Juan ___ un whisky.
Pablo ordered a beer and I.suspect that Juan a whisky
'Pablo ordered a beer and I suspect that Juan ordered a whisky.'

c. [Gapping, Embedding Factive]
Pablo pidió una cerveza y me molesta **que** Juan ___ un whisky.
Pablo ordered a beer and me bothers that Juan a whisky
'Pablo ordered a beer and I am bothered that Juan ordered a whisky.'

d. [No Gapping, No Embedding]
Pablo pidió una cerveza y Juan pidió un whisky.
Pablo ordered a beer and Juan ordered a whisky
'Pablo ordered a beer and Juan ordered a whisky.'

e. [No Gapping, Embedding Non-factive]
Pablo pidió una cerveza y sospecho **que** Juan pidió un whisky.
Pablo ordered a beer and I.suspect that Juan ordered a whisky
'Pablo ordered a beer and I suspect that Juan ordered a whisky.'

f. [No Gapping, Embedding Factive]
Pablo pidió una cerveza y me molesta **que** Juan pidió un whisky.
Pablo ordered a beer and me bothers that Juan ordered a whisky
'Pablo ordered a beer and I am bothered that Juan ordered a whisky.'

(Bîlbîie & de la Fuente 2019: 7)

The design of the second experiment was almost identical to that of the first. The subjunctive mood replaced the indicative mood in the non-elliptical items with emotive embedding predicates with the subjunctive, which is the preferred mood for these predicates. See (120) for an item from Experiment 2 in the crucial [No Gapping, Embedding Factive] condition with the true factive embedding verb *molesta* 'bothers'. As can be seen in (120), Bîlbîie & de la Fuente (2019) also replaced the introductory "wh-question [with] an initial locative/temporal adjunct that served the same function of contextual anchor" (Bîlbîie & de la Fuente 2019: 13). Lastly, there were changes to the number and inventory of the embedding verbs between experiments, see Table 2.3. While the changes in mood selection and anchoring play a role for Bîlbîie & de la Fuente's (2019) analysis (see Section 2.5.3), the reasons for the changes in the embedding verbs are not discussed.

(120) [No Gapping, Embedding Factive]
En el bar, Pablo pidió una cerveza y me molesta **que** Juan pidiera un whisky.
in the bar Pablo ordered a beer and me bothers that Juan ordered-SUBJ a whisky
'At the bar, Pablo ordered a beer and I am bothered that Juan ordered a whisky.' (Bîlbîie & de la Fuente 2019: 13)

For the sake of conciseness, I present the results of both experiments together (see Table 2.4). The two experiments use different scales: a 7-point Likert scale in Experiment 1 and a 10-point Likert scale in Experiment 2.[37] To facilitate the comparison between experiments, Bîlbîie & de la Fuente (2019) report *z*-transformed scores. I think absolute values are more informative because they provide a better impression of the scale. I therefore transformed the raw scores from Experiment 1 from a 7-point scale to a 10-point scale in Table 2.4.[38] Note that for the two factive sub-types in Experiment 1, Bîlbîie & de la Fuente (2019) only provide the overall mean values; they do not provide the numbers for the elliptical vs. non-elliptical variants. The reason for this is not discussed. The figure they provide on the matter (Bîlbîie & de la Fuente 2019: 11, Fig. 4) shows similar *z*-scores and overlapping error bars for the gapping and full clause conditions within both of the two types.

[37] Bîlbîie & de la Fuente's (2019) reasoning for using these different scales was that "Experiment 2 was part of a larger cross-linguistic study in which a 10-point scale was a more appropriate fit for the other languages studied" (Bîlbîie & de la Fuente 2019: 13, fn. 11). They do not elaborate on this point.

[38] I transformed the values by dividing them by 0.7, which yields a good approximation of what the values would have been on a 10-point scale. However, this transformation projects the 7-point scale on a scale 0-10 rather than 1-10, which causes some discrepancies. To illustrate: the non-transformed value for the true factives (bottom row) is 3.9, which is slightly below the mid point of a 7-point scale (4). The transformed value should equally be slightly below the midpoint of the new 10-point scale (5.5.). However, the transformed value is slightly above that point at 5.57. This is the largest discrepancy in the transformed data. My maths is not good enough to do this properly. Since the descrepancy is slight and Bîlbîie & de la Fuente (2019) mainly base their analysis on Experiment 2, I nevertheless think my transformation is sufficient.

Table 2.3: Embedding verbs of Experiments 1 and 2 (Bîlbîie & de la Fuente 2019: 8, 13). Differences between the experiments are highlighted in bold.

	Experiment 1	Experiment 2	Translation
Non factive	*creer*	*creer*	believe
($N = 8$)	*imaginarse*	*imaginarse*	imagine
	parecer	*parecer*	seem
	sospechar	*sospechar*	suspect
	suponer	*suponer*	suppose
	decir	*decir*	say
	rumoarse	*rumoarse*	be rumoured
	pensar	***contar***	think | tell
Semi-factive	*observar*	*observar*	observe
($N = 3 \mid 4$)	*saber*	*saber*	know
	ver	*ver*	see
		comprobar	confirm
True factive	*gustar*	*gustar*	like
($N = 5 \mid 6$)	*horrorizarse*	*horrorizarse*	be horrified
	molestarse	*molestarse*	be bothered
	sorprenderse	*sorprenderse*	be surprised
	impresionarse	***encantar***	be impressed | like
		preopcuparse	be worried

Table 2.4: Mean ratings per condition from Bîlbîie & de la Fuente's (2019: 9, 14) Experiments 1 (transformed from 7-point to 10-point scale) and 2 (original values).

	Experiment 1		Experiment 2	
	Gapping	Full clause	Gapping	Full clause
No embedding	9.36	9.44	9.46	9.16
Non-factive	8.97	8.87	8.98	9.10
Factive	6.66	6.59	7.44	8.30
Semi-factive	8.00		8.05	8.41
True factive	5.57		6.86	8.19

As for the statistical analysis, Bîlbîie & de la Fuente (2019) report that in both experiments, items without embedding receive significantly higher ratings than items with embedding. Within the items with embedding, non-factives receive significantly higher ratings than factives in both experiments. Within the factive predicates, semi-factives receive (descriptively) higher ratings than true factives in both experiments. As mentioned above, Bîlbîie & de la Fuente (2019) provide descriptive statistics only for the difference between factive sub-types.

The crucial difference between the experiments is that Experiment 2 featured significant interaction effects, while Experiment 1 featured no significant interaction effects. In Experiment 2, both the interaction of gapping and embedding on the one hand, and of gapping and factivity (i.e. factive vs. non-factive) on the other were significant. In items that featured clausal embedding, gapping received lower ratings than full clauses. Within these items, factives had a stronger negative effect on ratings than non-factives. Within the two factive sub-types ('true' factive vs. semi-factive), a visual inspection of the respective figure suggests the presence of an interaction as well (Bîlbîie & de la Fuente 2019: 17, Fig. 8): the negative effect of gapping is greater for 'true' factives than for semi-factives, for which the effect looks to be

very small. For the difference between factive sub-types, Bîlbîie & de la Fuente (2019) once more provide descriptive statistics only. Importantly, the negative effect of gapping is not so large as to cause outright ungrammaticality for any of the verb types. Participants made a statistically significant difference between these items and the unacceptable controls (Bîlbîie & de la Fuente 2019: 18, fn. 12).

The absence of any significant interactions in Experiment 1 is unexpected. Bîlbîie & de la Fuente (2019) identify the [No Gapping, Embedding Factive] condition as the cause for this. Ratings for this condition were lower than expected, given that there is no obvious structural ungrammaticality in these items. Bîlbîie & de la Fuente (2019) adduce two possible explanations for the low ratings: mood selection and the presence of the introductory anchor-question. The fact that the embedded clause invariably contained verbs in the indicative may have lowered ratings in this condition, given that the subjunctive is preferred over the indicative for emotive verbs. For the second possible reason, the presence of the anchor question, they write:

> "Note that the second conjunct introduced by a factive verb does not answer that question [i.e, the provided anchor-question preceding the items]; rather it addresses a different QUD (= Question Under Discussion), e.g. no longer what someone did, but what effect it had on the speaker, leading to a discourse incoherence." (Bîlbîie & de la Fuente 2019: 12)

Recall that both these potentially relevant properties (mood selection and anchor-question) of the experimental design were changed between Experiments 1 and 2. By implementing these two changes, both potential causes for the unexpectedly low ratings in the [No Gapping, Embedding Factive] condition were manipulated simultaneously. As a consequence, it is impossible to deduce which of the two was crucial for the interaction effects in Experiment 2. It is therefore surprising that Bîlbîie & de la Fuente (2019) attribute the difference between the experiments exclusively to the change in mood selection. They write:

> "[I]t seems reasonable to assume that the potential confound in our Experiment 1 (i.e. mood selection) did have an effect on the results. The mood selection created a floor effect in our Experiment 1, as the indicative choice (the marked form compared to the more natural subjunctive) cancelled the effect of our experimental factors. In particular, this choice seems to have been responsible for the lack of interaction between Gapping and Embedding, and between Gapping and Factivity in Experiment 1." (Bîlbîie & de la Fuente 2019: 17 f.)

There is no mention of the introductory *wh*-question of Experiment 1. Also, it is unclear what is meant by "floor effect". The means of the relevant conditions do not approach the lowest possible rating (the "floor"), the lowest of the observed means being around the mid point of the scale (5.57 for true factives in Experiment 1, see Table 2.4). In their analysis, Bîlbîie & de la Fuente (2019) presuppose that mood selection accounts for the absence of interaction effects in Experiment 1, and that the data from Experiment 2 with these interaction effects represents the true pattern in locally embedded gapping. In the following section, I offer a critical summary of their analysis.

2.5.3 A critical discussion of Bîlbîie & de la Fuente's (2019) analysis

The crucial observations from Bîlbîie & de la Fuente's (2019) acceptability study are summarized in (121). Observation 1, that locally embedded gapping is indeed acceptable in Spanish, is probably the most crucial one. It follows straightforwardly from Bîlbîie et al.'s (2021) results and is in line with the theoretical literature on Spanish (see above). In their analysis, Bîlbîie & de la Fuente (2019) focus on explaining Observations 2 to 4, which I summarize and offer criticism on in this section.

(121) **Observations on Spanish locally embedded gapping** (Bîlbîie & de la Fuente 2019: 18)

1. Locally embedded gapping (with the complementizer *que*) is acceptable in Spanish.
2. There is a hierarchy of the types of embedding predicates in terms of acceptability: Non-factives > Semi-factives > True factives.
3. The verb type effect does not cause outright ungrammaticality for any verb type.
4. The verb type effect is specific to ellipsis.

Observation 2 is problematic. It is unclear whether it strictly follows from either Bîlbîie & de la Fuente's (2019) data or their analysis. As outlined above, the effect of gapping for factives seems to be driven by the true factives. The effect within semi-factives seems to be very small, resembling the effect within the non-factives. However, Bîlbîie & de la Fuente (2019) do not conduct the necessary statistical tests for the two sub-types of factives to show that the difference is significant. Furthermore, in their analysis (see below), Bîlbîie & de la Fuente (2019) treat non-factives and semi-factives alike in some crucial semantic/pragmatic aspects (see below). However, in another part of their paper they explicitly speak of a "continuum [they] observed along three different semantic classes (namely, non-factive, semi-factive and true factive)" (Bîlbîie & de la Fuente 2019: 28). In the greater context of this quote, it is reasonable to interpret this as to refer to the interaction of gapping and verb type (rather than to the effect of verb type alone).

In addition to these statistical concerns regarding the verb types, the experimental design does not seem to be ideal to find meaningful differences between the two factive sub-types in the first place. The number of embedding predicates is not balanced and overall low (8 non-factives, 4 semi-factives and 6 true factives in Experiment 2 – see Table 2.3 in Section 2.5.1 above). The number of embedding verbs (especially that of the semi-factives) is thus probably too low to tease apart the individual effects of the verbs from the general type these verbs represent. Notice that the uneven number of verbs means that some of them feature in more experimental items than others (Bîlbîie & de la Fuente 2019: 13), which creates an imbalance in design. I will still follow Bîlbîie & de la Fuente (2019) in assuming their Observation 2, not least because it opens up an interesting avenue in the investigation of *que* in globally embedded gapping (Chapter 3). However, Observation 2 should be treated with caution.

To account for the observations in (121), Bîlbîie & de la Fuente (2019: 28) propose an analysis based on the semantic and discourse properties of gapping on the one hand, and of the different verb types on the other hand. They assume this is generally more suitable than a syntactic analysis where complement clauses to factives and non-factives differ in structure (such as Fernández-Sánchez 2021). According to Bîlbîie & de la Fuente (2019: 28), syntactic constraints cannot account for the gradual difference between verb types in their data because analyses that work with such constraints predict categorical differences

(grammatical vs. ungrammatical). They argue that one would require a third structure for semi-factive complement clauses in order to model the difference between the factive sub-types (Bîlbîie & de la Fuente 2019: 23). I will argue against this view in Chapter 3.5 based on my data on globally embedded gapping.

Regarding the discourse properties of clause-embedding predicates, Bîlbîie & de la Fuente (2019: 31 f.) follow Farkas (2003) in assuming that non-factives and semi-factives both have the property of being "assertive". This "assertive" property contrasts these two verb types from true factives, which are "non-assertive". It is not entirely clear what they mean by this; they provide circular explanations such as "their complements have assertive CCP [Context Change Potential]" (Bîlbîie & de la Fuente 2019: 29) and "these two kinds of predicates […] are assertively added to an epistemic context" (ibid.). I assume they mean that *the propositional content of the complement clause to* these two kinds of predicates (non-factives and semi-factives) are added to an epistemic context, rather than the predicates themselves.[39] Bîlbîie & de la Fuente (2019) also make the classic assumption that (both sub-types of) factives presuppose the content of their complement. The resulting classification of embedding verb types is summarized in Table 2.5. As Bîlbîie & de la Fuente (2019) speculate, the degradation for true factives may arise from a feature mismatch between the conjuncts. The antecedent is not embedded and therefore [+ assertive], which matches with the [+ assertive] feature of the gapped clause under non-factives and semi-factives. Under true factives, the gapped clause is [- assertive], thus not matching the antecedent.

Table 2.5: The featural make-up of the three types of embedding predicates (Bîlbîie & de la Fuente 2019: 30)

	Assertive	Presupposition trigger
Non-factive	+	−
Semi-factive	+	+
True factive	−	+

There are some problems with this approach. First, it conflates non-factives and semi-factives in the relevant feature [assertive]. The differences between the two verb types observed in the experiments is thus unexpected. Second, it makes use of binary features, which predict binary judgments. It is unclear why a mismatch of features would lead to a degradation rather than outright unacceptability. Bîlbîie & de la Fuente (2019) try to ameliorate this by appealing to the gradualness of assertion: "the more asserted the complement is, the easier it is embedded" (Bîlbîie & de la Fuente 2019: 30). While gradability of assertion is something that is discussed in the literature (Wolf 2014, Greenberg & Wolf 2018), it is unclear how this would apply to the case of locally embedded gapping.

The third problem with this aspect of the analysis is that it is not specific to gapping, but should apply to non-elliptical sentences as well. The (apparent) interaction within true factives is without explanation at this point. Bîlbîie & de la Fuente (2019) address the issue of ellipsis-specificity in the second part of their analysis, which is based on discourse restrictions on gapping. Earlier research on gapping (Kuno 1976, Steedman 1990, Reich 2007, Toosarvandani 2016 amongst others) observed that in gapping, the two conjuncts must generally refer to a joint discourse topic (more so than non-elliptical coordinate clauses).

[39] I will go into detail on what I understand under the term 'assertion' in Chapter 3.1.1. My definition of 'assertion' may actually be identical to what Bîlbîie & de la Fuente (2019) had in mind in their use of the term, but I am not able to tell for sure due to the vagueness of their description. In Chapter 3.1.1, I also provide definitions for other technical terms I use in my discussion (including 'presupposition' and 'factivity'). For now, I will continue under Bîlbîie & de la Fuente's (2019) use of terminology.

Repp (2009) formulates this observation in her "Principle of Balanced Contrast" in (122). Johnson (2018) puts the same observation in terms of the notion of "Question Under Discussion'" (QUD, Roberts 1996, 2012) in his constraint in (123). Bîlbîie & de la Fuente (2019) center their discussion around this QUD-approach.

(122) **Principle of Balanced Contrast, Conjunction 'and'** (Repp 2009: 83)
In gapping, both conjuncts must make the same kind of contribution to a common discourse topic.

(123) **QUD constraint on gapping** (Johnson 2018: 601)
XP can gap or strip if it is proffered as a partial answer to the Question Under Discussion, Q, and XP's meaning can be recovered from Q.

According to Bîlbîie & de la Fuente (2019), this pragmatic property of gapping accounts for the degradation of locally embedded gapping under factives. When the gapped clause is embedded under a factive predicate, the content of the complement clause is backgrounded in discourse (Anand & Hacquard 2014) and can therefore not provide an answer to the same QUD as the non-embedded antecedent clause. Under non-factives, the same backgrounding does not happen. Complement clauses to non-factives thus can provide an answer to the same QUD as the antecedent clause.

Let me illustrate this with an example. Imagine that English (like Spanish) allowed locally embedded gapping with a complementizer. The non-embedded gapping example in (124b) and locally embedded gapping under a non-factive like *think* in (124c) would be appropriate utterances, because both conjuncts provide (partial) answers to the QUD in (124a). Locally embedded gapping under a factive such as *regret* in (124d) would not be an appropriate utterance, because only the first conjunct answers the QUD in (124a). In the second conjunct, the main point lies with the content of the matrix clause (i.e. the attitude of the first person matrix subject), and not the content of the embedded clause.

(124) a. **QUD**
Who drinks what for breakfast?

b. **Non-embedded gapping**
John drinks coffee and Mary ___ tea.

c. **Locally embedded gapping under a non-factive verb (Pseudo-English)**
John drinks coffee and I think **that** Mary ___ tea.

d. **Locally embedded gapping under a factive verb (Pseudo-English)**
#John drinks coffee and I regret **that** Mary ___ tea.

In general, the individual contribution of each of the two central aspects of Bîlbîie & de la Fuente's (2019) analysis (i.e. assertiveness and QUD-sensitivity) is unclear. There seems to be a tight connection between the two: in order to have the potential to address a QUD, a proposition must be asserted. As for the difference between semi-factives and true factives, Bîlbîie & de la Fuente (2019) make reference to an observation by Anand & Hacquard (2014) that "occasionally semi-factives (such as *know*) can behave like non-factives, i.e. the content of its embedded complement becomes main point" (Bîlbîie & de la Fuente 2019: 30). This means that the complement of semi-factives can sometimes address a QUD and thus be acceptable in gapping, while the complement of true factives (presumably) can never address a QUD. It

remains unclear what determines this variable behaviour of semi-factives, i.e. under what circumstances they can or cannot address a QUD.

In addition to the potential to address a QUD, Bîlbîie & de la Fuente (2019) point out another way that the backgrounding associated with factive complements interacts with gapping. The constituents in a gapping clause are contrastively focussed, and therefore have a prominent status in discourse (see also Bîlbîie et al. 2023: 32). Such prominent elements are infelicitous in an environment that is backgrounded in discourse such as the complement to a factive verb. Bîlbîie & de la Fuente (2019: 32) assume that this infelicity further contributes to the decrease in acceptability for factives. With this discourse-prominence of the remnant constituents, Bîlbîie & de la Fuente (2019) add a third potential explanation for the degradation under factives to the already proposed assertion and QUD-related explanations. Once more, it is unclear how discourse-prominence interacts with assertiveness and QUD-sensitivity.

The gradualness in Bîlbîie & de la Fuente's (2019) data, especially the intermediate position of semi-factives between non-factives and true factives, raises more general problems. Bîlbîie & de la Fuente (2019) claim that their semantic/discourse based approach is suitable to model this gradualness. However, they conceptualize their two central aspects in terms of binary properties: a verb is either [+ assertive] or [- assertive], and the complement of a verb is either fore- or backgrounded. For non-factives and true factives, these properties seem to be set invariably; semi-factives, on the other hand, seem to be able to switch more or less freely between these properties, which leads to their middle position in terms of mean acceptability ratings.

Crucially, this assumed variability does not put semi-factives in a gradual middle position between non-factives and true factives from a conceptual perspective. Instead, semi-factives are expected to categorically pattern with non-factives or categorically pattern with true factives in any individual instance in which they are interpreted. In the context of Bîlbîie & de la Fuente's (2019) experiments, this means that each time a participant reads an item and judges it in the experiments, the verb should receive an interpretation that is either non-factive (i.e. assertive with foregrounded complement clause) or factive (i.e. non-assertive, with backgrounded complement). A truly intermediate interpretation should not be possible.

This leads to the question: if semi-factives can be (freely?) interpreted as non-factives, why do they not consistently pattern as such when the construction they occur in pushes them to? In other words, why is there a difference between non-factives and semi-factives in locally embedded gapping at all? One would expect participants to consistently accommodate the non-factive interpretation for semi-factives, given that this would be the only interpretation appropriate in locally embedded gapping. Acceptability ratings should thus not be lower for semi-factives than for non-factives. Further, it would be expected that the categorical interpretation shows in the individual ratings in form of a bimodal distribution. If participants interpret semi-factives as non-factives, ratings should be high. If participants interpret semi-factives as true factives, ratings should be low. Individual ratings should, on the whole, not be intermediate.

It may very well be the case that there is such a bimodal distribution, which leads to the intermediate mean ratings Bîlbîie & de la Fuente (2019) observe for semi-factives. Unfortunately, they do not discuss the distribution of the individual ratings. Neither do they report the standard deviations for the factive sub-types, which would have provided an impression of the distribution of the data around the mean.

The standard deviation would be larger for semi-factives than for the other verbs, if the ratings for semi-factives had indeed been distributed bimodally.

One aspect that remains implicit in Bîlbîie & de la Fuente's (2019) analysis concerns QUD-accommodation in non-elliptical sentences. As discussed above, part of their analysis is based on restrictions on the QUD that gapping relates to. However, they do not discuss QUD-accommodation in full clauses. This aspect of QUD-accommodation is especially relevant for embedding under true factives because this verb type shows a comparatively large negative effect of gapping. As discussed above, Bîlbîie & de la Fuente (2019) account for this effect by assuming that the backgrounding under true factives makes it so that the elliptical conjunct cannot be used to answer the same QUD as the antecedent clause. However, this backgrounding is an effect of the verb, not of gapping, and thus obtains in full clauses as well. The example in (125) shows that for a QUD that licenses non-embedded gapping, local embedding under a true factive is infelicitous, regardless of whether the embedded clause is elliptical (125a) or not (125b).

(125) **QUD**
Who drinks what for breakfast?

a. **Gapping (Pseudo-English)**
#John drinks coffee and I regret **that** Mary ___ tea.

b. **Full clause**
#John drinks coffee and I regret **that** Mary drinks tea.

To account for the fact that gapped clauses under true factives receive lower ratings than full clauses, Bîlbîie & de la Fuente (2019) should have discussed this aspect of QUD-accommodation. I will offer a short possible route such a discussion could take. For this, I make the (likely uncontroversial) assumption that QUD-accommodation for non-elliptical coordinate clauses is not subject to the same restrictions that gapping is but instead proceeds more liberally. This means that the set of possible QUDs to be accommodated is larger in number and more varied in syntactic form for coordinate full clauses than for coordinate clauses with gapping.

As a consequence of this assumption, the effect within true factives may be due to a different QUD that is accommodated in full clauses than in gapped clauses. What possibly happens in coordinate full clauses is that rather than a singular QUD that applies to both conjuncts, each clause can accommodate its own QUD. The non-elliptical sentence in (127) would then in principle be felicitous because the accommodation of a coordinate QUD like (126) is possible, where each conjunct of the coordinate structure (127) targets one question of the coordinate QUD. This mechanism of QUD-accomodation is speculative; more research is required, especially given the fact that the discourse in (126) and (127) does not seem very natural (albeit not outright infelicitous).

(126) **QUD**
What does John drink and how do you feel about Mary's drinking tea?

(127) John drinks coffee and I regret **that** Mary drinks tea.

Still, the fact that Bîlbîie & de la Fuente's (2019) analysis in general relies so strongly on the notion of QUD brings us back to one of the possible confounds of Experiment 1: the introductory anchor question.

The example in (129) repeats the sample item from (120) above, in the [No Gapping, Embedding Factive] condition with the true factive embedding verb *molesta* 'bother'. The *wh*-question in (128) preceded this item in the questionnaire. It simultaneously applies to the first conjunct and the embedded clause in the second conjunct and thus licenses gapping. Recall that the results of Experiment 1 indicated no meaningful difference between gapping and non-gapping items embedded under true factives. I suggest that this lack of an effect follows from the fact that the anchor question prevented participants from freely accommodating any QUD. Instead, with the provided anchor question setting the QUD, it is to be expected that an answer featuring embedding with a true factive will be infelicitous, regardless of ellipsis. It is therefore likely that the presence of the anchor question is responsible for the lack of a gapping effect for true factives in Experiment 1. This goes against Bîlbîie & de la Fuente's (2019) explanation in terms of a floor effect on ratings caused by mood choice.

(128) ¿Qué bebidas pidieron los dos amigos en el bar?
what drinks ordered the two friends in the bar
'What drinks did the two friends order in the bar?'

(129) [No Gapping, Embedding Factive]
Pablo pidió una cerveza y me molesta **que** Juan pidió un whisky.
Pablo ordered a beer and me bothers that Juan ordered-IND a whisky
'Pablo ordered a beer and I am bothered that Juan ordered a whisky.' (Bîlbîie & de la Fuente 2019: 7, reproduced from 120 above)

It is unfortunate that Bîlbîie & de la Fuente (2019) rely on a theoretical discussion on QUDs, given that their experimental materials would have made it possible to investigate the relevance of an explicit QUD experimentally. They could have conducted an additional experiment that checked the acceptability for items with vs. without the explicit QUD. If items with the introductory QUD such as (128) returned lower ratings, this would be good evidence that QUD-accommodation indeed plays a role.

In this section, I have outlined Bîlbîie & de la Fuente's (2019) analysis of the verb type effects they find for locally embedded gapping. My critical examination has revealed some shortcomings in their analysis. In the following section, I bring the discussion back to globally embedded gapping. I relate Bîlbîie & de la Fuente's (2019) findings and analysis to the Typological Conjecture, especially with respect to predictions regarding constraints on the complementizer in globally embedded gapping.

2.5.4 Relevance of Bîlbîie & de la Fuente's (2019) results and analysis for globally embedded gapping

I summarize the possible causes for the verb type effects in Bîlbîie & de la Fuente's (2019) experiments in (130). In the preceding section, I outline a number of problematic points regarding these potential causes, especially in regard to the possible interconnection of causes 2 to 4. In this section, I will turn back to the comparison of locally vs. globally embedded gapping from the perspective of the Typological Conjecture. I reproduce my conjecture in (131) from (97) in Section 2.3 above. The conjecture predicts that the verb type restrictions that Bîlbîie & de la Fuente (2019) observe for locally embedded gapping, also operate on *que* in globally embedded gapping.

(130) **Possible causes for the verb type effects in Bîlbîie & de la Fuente (2019)**

1. Mood mismatch
2. Assertiveness
3. QUD-sensitivity
4. Discourse prominent constituents within the elliptical clause

(131) **Typological Conjecture on embedded gapping**
Any constraint on a complementizer specific to ellipsis that operates in local embedding also operates in global embedding (and vice versa).

In marked contrast to the Typological Conjecture, none of the potential causes for the verb type effects in (130) predict the same verb type effects for *que* in globally embedded gapping. Causes 1 to 3 relate to mismatches between the conjuncts that do not obtain in globally embedded gapping. See (132) for one of the corpus examples initially discussed in (106) above, in which both of the coordinate clauses are embedded under the verb *ocasiona* 'cause'. When the whole coordinate structure is embedded like this, both conjuncts are equal on a syntactic and semantic/pragmatic level. There is no opportunity for a mismatch to arise from the way co- and subordination interact in globally embedded gapping. This absence of a mismatch between conjuncts is independent of the presence or absence of the complementizer in the second conjunct.

(132) El sistema ocasiona que las cervazas más baratas paguen más impuestos y (**que**) las
the system causes that the beers most cheap pay-SUBJ more taxes and that the
mas caras ___ menos.
most expensive less
'The system results in the cheapest beers paying more tax and the most expensive ones paying less tax.' (corpus example, repeated from 106 above)

Causes 1 to 3 therefore predict no *que* effect and no verb type effect. Regarding cause 1, the embedded verbs in both conjuncts match in mood since the same predicate selects both conjuncts. See (133) for a non-elliptical version of (132), where the embedding predicate *ocasiona* 'cause' selects for the subjunctive form of the embedded verb *paguen* 'pay' in both conjuncts. While other embedding predicates may select for the other mood (i.e. the indicative), no embedding predicate is such that it selects a different mood in each conjunct. Furthermore, mood selection is independent of the complementizer, so *que* cannot lead to a mismatch either. The same reasoning goes for Cause 2 since the assertive status of the propositions is identical in the two conjuncts (assuming with Bîlbîie & de la Fuente 2019: 29 that this status is determined by the embedding predicate).

(133) El sistema ocasiona que las cervazas más baratas paguen más impuestos y **que** las mas
the system causes that the beers most cheap pay-SUBJ more taxes and that the most
caras paguen menos.
expensive pay-SUBJ less
'The system results in the cheapest beers paying more tax and the most expensive ones paying less tax.' (non-elliptical variant of corpus example 106 above)

As for Cause 3, once more both conjuncts behave identically. They can both serve as partial answers to the same QUD, irrespective of the presence of *que* in the second conjunct, and irrespective of the factivity

of the embedding verb. Any QUD that can be accommodated for an embedded gapping structure with *que* will also be one that can be accommodated for the same structure without *que*. This is illustrated in (134) and (135): on hearing the sentence in (135), an interlocutor can plausibly accommodate the QUD in (134), regardless of the presence of *que* in the second conjunct. The sensitivity to factivity is unexpected under a QUD approach because the embedding predicate stands outside the conjunction and therefore should have no influence on properties of interpretation specific to gapping, such as QUD-accommodation.

(134) **QUD**
What does the system cause as to what beers pay how much tax?

(135) El sistema ocasiona que las cervazas más baratas paguen más impuestos y **que** las mas
the system causes that the beers most cheap pay-SUBJ more taxes and that the most
caras ___ menos.
expensive less
'The system results in the cheapest beers paying more tax and the most expensive ones paying less tax.' (corpus example, repeated from 106 above)

Cause 4 relates to the observation that the constituents in the gapping conjunct have a discourse-prominent status because they are contrastive foci (Bîlbîie & de la Fuente 2019: 32). For locally embedded gapping, Bîlbîie & de la Fuente (2019) argue that this discourse prominence leads to a pragmatic incongruity with the backgrounding associated with the complement of factive predicates, and thus decreases acceptability ratings. For globally embedded gapping, the same effect is to be expected because the discourse prominence of the constituents in the elliptical clause should equally clash with the backgrounding of factive predicates. This clash should obtain independently of whether the clause hosting the foci is part of a coordinate structure or not. Cause 4 therefore predicts a general verb type effect in globally embedded gapping, with ratings on the whole lower for factives than non-factives. However, it predicts no main or interaction effect of *que*, since the discourse-prominent status of the constituents in the elliptical conjunct is independent of the complementizer.

Summing up, the different aspects of Bîlbîie & de la Fuente's (2019) analysis of locally embedded gapping make fundamentally different predictions as to the acceptability of *que* in globally embedded gapping. There should be no verb type effect on *que* in globally embedded gapping that is specific to ellipsis. The absence of this prediction is due to the mismatch of conjuncts being the central piece to their analysis. The Typological Conjecture, on the other hand, predicts that *que* underlies the same verb type effects in globally embedded gapping, and that these effects are specific to ellipsis. For Bîlbîie & de la Fuente's (2019) account it is of course, not a problem since they are not concerned with global embedding.

One more prediction Bîlbîie & de la Fuente's (2019) account makes concerns the cross-linguistic behaviour of locally embedded gapping. Let's assume that the factive vs. non-factive distinction exists for clause-embedding predicates in other languages, and the aspect of QUD-sensitivity likewise exists for gapping in other languages. (Both of these assumptions do not seem controversial.) It follows, then, that the sensitivity of locally embedded gapping to the type of embedding predicate should be cross-linguistically robust. In other words, we expect to find that other languages display a degradation of locally embedded gapping under factives like Spanish does. Furthermore, if we follow Bîlbîie & de la Fuente's (2019) analysis, we expect that within factives, there should be a difference between semi-factives and true fac-

tives. The following section discusses evidence from English, Farsi, Romanian and Dutch, which strongly suggests that this cross-linguistic prediction is borne out.

2.6 Locally embedded gapping and factivity in other languages

In Section 2.1 I discussed the dependency of locally embedded gapping in English and German and the presence of the complementizer. The claim in the literature (Weir 2014, Reich 2007) is that with the complementizer, locally embedded gapping is unacceptable in these languages, but without the complementizer, acceptability is much better. Put differently, if there is an acceptable version of locally embedded gapping, it is the one without the complementizer.

Not all judgments in the literature agree with this assessment. Sag (1977), Hankamer (1979) and Johnson (1996/2004) consider locally embedded gapping ungrammatical, regardless of whether the complementizer is absent or present, see (136) and (137). Johnson softens this judgment in later work (Johnson 2018) to accommodate Weir's (2014) observations on the relevance of the complementizer. For a more detailed description of these approaches, see Section 2.2.

(136) * Alan went to New York, and {I know / it seems} (**that**) Betsy ___ to Boston. (Sag 1977: 139)

(137) * Alfonse stole the emeralds, and I think (**that**) Mugsy ___ the pearls. (Hankamer 1979: 19)

Even without the complementizer, the acceptability of locally embedded gapping seems to be limited. Weir (2014) provides a number of examples, see (138), but is hesitant to mark many of them as fully acceptable. Instead, he reports nuanced introspective judgments. According to Weir (2014), this nuance in the judgments in (138) seems to be caused by the embedding verb: 'bridge verbs' (Erteschik-Shir 1977) seem to allow locally embedded gapping more easily than, e.g. 'response stance' predicates (e.g. *agree, accept, deny*, Weir 2014: 234) or factive predicates (e.g. *know, find out, be surprised*, Weir 2014: 234). However, Weir (2014) does not mark all his examples with bridge verbs as fully acceptable either, which adds to the empirical puzzle. The crucial conclusion for my purposes is the obvious parallel to Bîlbîie & de la Fuente's (2019) findings on Spanish: the type of embedding predicate seems to impact acceptability; specifically, factive predicates do not seem to allow locally embedded gapping.

(138) John ate oysters…

a. and I {?think/?believe/??hope/suspect/?was told/imagine} Mary ___ swordfish.

b. and I {?*found out/*remember/*deny/?*know} Mary ___ swordfish.

c. and I {*am proud/*am angry/*am surprised} Mary ___ swordfish. (Weir 2014: 333)

To clear up this difficult empirical situation Bîlbîie et al. (2023) conducted an acceptability study in English. The study contained two experiments. One experiment tested *that*-omission in gapping vs. full clauses under non-factive verbs ($N = 10$). The other experiment did the same under factive verbs. Like in the Spanish experiments in Bîlbîie & de la Fuente (2019) (see Sections 2.5.2 and 2.5.3), Bîlbîie et al. (2023) split the factive predicates into semi-factive predicates ($N = 5$) and true factive predicates ($N = 5$). All in all, the experimental results show that the omission of *that* improves acceptability ratings in gapping but not in full clauses. This effect is largest for non-factives. For both factive sub-types, gapping receives

ratings within the unacceptable range (partially worse ratings than unacceptable controls). Within this unacceptable range, however, *that*-omission likewise seems to display an ameliorating effect, albeit only for semi-factives. For true factives, *that*-omission makes no difference in gapping, which is likely due to the fact that true factives do not allow for *that*-omission in the first place (which Bîlbîie et al. 2023 find out in another experiment).

Bîlbîie et al.'s (2023: 413) interpretation of the results is that in terms of factivity, English shows the same gradual three-way distinction like Spanish (non-factive > semi-factive > true factive). While this may be true for the overall ratings for the three verb types (i.e. across all conditions, both elliptical and non-elliptical as well as with and without *that*), it is unclear how this applies to either gapping, or the interaction of gapping and *that*-omission. Experiment 2 finds no significant interaction of gapping and factive sub-type (Bîlbîie et al. 2023: 410), which suggests that semi-factives and true factives react identically to gapping generally. As for the interaction of *that*-omission and gapping, non-factives and semi-factives seem to behave alike in their respective experiments. Rather than a middle position between non-factives and true factives, semi-factives thus display a variable behaviour, which classes them with non-factives in one regard (i.e. *that*-omission) but with true factives in another (i.e. licensing locally embedded gapping).

Whatever the nuanced details regarding the factive sub-types might be, the relevant result of Bîlbîie et al.'s (2023) experiments is that the type of embedding verb affects the local embeddability of gapping in English (as predicted by Weir 2014). Additionally, the relevant verb type, on the whole, seems to coincide with the non-factive vs. factive distinction. This clearly parallels Bîlbîie & de la Fuente's (2019) results on Spanish and is present in other languages as well.

In an acceptability study on gapping in Farsi, Bîlbîie & Faghiri (2022) find the same sensitivity to verb type. The experimental design in Bîlbîie & Faghiri (2022) is the same as that for Spanish in Bîlbîie & de la Fuente (2019). The results by-and-large parallel those of Bîlbîie & de la Fuente (2019): gapping is degraded with factives as opposed to non-factives, and within factives, true factives show a larger negative effect of gapping than semi-factives. An acceptability study on Romanian and French (Bîlbîie et al. 2021) seems to show the same verb type effects in these languages (at least upon visual inspection of the plots and tables).[40]

Finally for Dutch, a language that behaves similarly to English as regards complementizer optionality (see above), Broekhuis & Corver (2019) report the introspective judgments that locally embedded gapping is ungrammatical with embedding *weet* 'know', but "much better" (Broekhuis & Corver 2019: 286) with *denk* 'think', see (139). They, too, assume that the relevant difference between *denk* and *weet* is factivity.

(139) **Dutch, locally embedded gapping**
Jan leest het Artikel en ik {*weet / denk} Marie ___ het boek.
Jan reads the article and I know think Marie the book
'Jan is reading the article and I know/think Marie is reading the book.' (adapted from Broekhuis & Corver 2019: 286 f.)

In sum, the sensitivity of locally embedded gapping to whether or not the embedding verb is factive seems to be relevant not only in Spanish, but cross-linguistically. In fact, all languages that have undergone any study display such sensitivity. Notice that these languages are heterogeneous as to whether they

[40] Bîlbîie et al. (2021) is in French, which means that I can't read it.

allow complementizers in globally embedded gapping: in Spanish and Romanian, the complementizer is acceptable, while in English, French, Dutch and Farsi, the complementizer is ungrammatical. (Farsi is somewhat of a special case, see Section 2.3 for discussion.) This indicates that the observed verb type effects in locally embedded gapping are not related to the complementizer as such. This observation generates a prediction that runs counter the predictions from the Typological Conjecture: there should b no verb type constraints on the complementizer in globally embedded gapping.

2.7 Summary and predictions for Spanish globally embedded gapping

I started this chapter with an outline of the patterns of English and German complementizers in two configurations of embedded gapping, which I call "global" and "local" embedding. The facts regarding the complementizers in these two languages are well-known but seldom jointly examined: In neither globally nor locally embedded gapping can the elliptical conjunct host a complementizer. This is notably different from full clauses, in which the complementizer is either generally optional (in global embedding) or subject to constraints driven by the embedding predicate (in local embedding). In Section 2.2, I looked at some of the more prominent theoretical approaches to (English) gapping. While locally or globally embedded gapping play roles of varying importance within (some of) these approaches, none of them feature any in-depth treatment of both embedded gapping configurations. In some approaches, this lacking treatment leads to false predictions of the acceptability of the complementizer in English and, by extension, to German.

In Section 2.3, I took a step back from theoretic questions, and widened the empirical scope to languages beyond English and German. I examined the available literature for a number of languages regarding the acceptability of complementizers in the two structures of embedded gapping. I showed that while (unsurprisingly) not all languages behave like English and German, there seems to be a cross-linguistic pattern for complementizers in embedded gapping: ellipsis-specific constraints on complementizers within languages are identical in locally and globally embedded gapping. I formulate this pattern as a "Typological Conjecture" in (93), which is reproduced here in (140).

(140) **Typological Conjecture on embedded gapping**
Any constraint on a complementizer specific to ellipsis that operates in local embedding also operates in global embedding (and vice versa).

In Section 2.4, I investigated whether Russian may provide evidence for the Typological Conjecture. I examined the available literature and presented new experimental data on locally embedded gapping in Russian, which I compared to my experimental findings on globally embedded gapping in Bonke (forthcoming). I found that the availability of locally embedded gapping in general and the availability of the complementizer *čto that* are subject to the same constraint on the discourse relation between the conjuncts. In contrastive coordination, locally embedded gapping and *čto* in globally embedded gapping receive degraded ratings. In parallel, non-contrastive coordination, I found no such degradation. However, these results are not suitable to count as evidence in favour of the Typological Conjecture, since the observed effects are not specific to ellipsis, but surface in non-elliptical sentences as well.

In Section 2.5, I turned to Spanish. In this language, the availability of locally embedded gapping is subject to constraints on the embedding predicates. I discuss the paper by Bîlbîie & de la Fuente (2019) in detail, who argue that locally embedded gapping is constrained by the type of embedding verb. They conducted two acceptability judgment experiments that tested non-factive embedding verbs against two sub-types of factives, namely semi-factives (= verbs of knowledge and perception) and 'true' factives (= emotive verbs). In their findings, locally embedded gapping receives relatively high acceptability ratings with non-factives, slightly degraded ratings with semi-factives, and more strongly degraded ratings with 'true' factives.

These findings provide an ideal testing ground for the Typological Conjecture, which straightforwardly predicts the same factivity constraints on the complementizer *que* 'that' in globally embedded gapping. However, Bîlbîie & de la Fuente's (2019) analysis of their locally embedded gapping data is based on mismatches between the conjuncts, which (in a nutshell) predicts no such effects. This prediction of no effects is supported by the observation in Section 2.6 that factivity driven restrictions on locally embedded gapping are not specific to Spanish, but surface in a number of languages. These languages are heterogeneous as to whether they allow for complementizers in globally embedded gapping, which suggests that what Bîlbîie & de la Fuente (2019) observe for Spanish is not tied to the language-specific behaviour of the complementizer, which is a core property of the Typological Conjecture. Instead, the cross-linguistic pattern of verb-type effects on locally embedded gapping may arise from the semantic/pragmatic differences between types of clause-embedding verbs on the one hand, and discourse properties of gapping on the other hand, both of which seem to surface in many languages.

Thus, the predictions of the Typological Conjecture and Bîlbîie & de la Fuente's (2019) analysis of locally embedded gapping are radically different. Since I firmly believe the proverbial proof to be in the equally proverbial pudding, I test whether *que* in globally embedded gapping is sensitive to the type of embedding predicate in the following chapter.

3 The complementizer *que* 'that' in Spanish embedded gapping

In this chapter, I present the results of an acceptability study on the complementizer *que* 'that' in Spanish embedded gapping. Specifically, I investigated *que* in the elliptical conjunct in sentences like the one in (1). I call this structure "globally embedded gapping"; see Chapter 2 for a discussion of this structure and choice of terminology. Parts of this chapter have previously been published in Bonke & Repp (2022).

(1) El[sic] cree que el mundo es su empresa y **que** los mexicanos ___ sus lacayos.
He thinks that the world is his company and that the Mexicans his lackeys
'He thinks that the world is his company and that the Mexicans are his lackeys.' (corpus example from Bonke & Repp 2022: 528, reproduced from example 87 in Chapter 2.3)

The list in (2) summarizes the hypotheses I derived in Chapter 2 on how the acceptability of *que* may be constrained in globally embedded gapping. Hypothesis 1 formulates the potential influence of the type of the embedding verb in this regard. This hypothesis is the most important one, which is why I will return to it in more detail shortly.[1] Next to this first hypothesis, I investigate whether there is a difference between gapping and stripping (Hypothesis 2), and whether the patterns I observe within gapping are specific to (this type of) ellipsis, or whether they instead represent general patterns of coordinate clauses (Hypothesis 3). Despite the focus on these specific hypotheses, I would like to point out that this chapter first and foremost represents a general contribution to the study of globally embedded gapping, and the role of the complementizer within that structure – a topic which has received almost no attention in ellipsis research.

(2) **Hypotheses**: The acceptability of *que* in global embedding is constrained by
1. the type of embedding verb (non-factive vs. semi-factive vs. true factive)
2. the ellipsis type (gapping vs. stripping)
3. ellipsis as such (gapping vs. full clauses)

Regarding Hypothesis 1, there are specific predictions for different types of embedding verbs. In short, I hypothesize that the acceptability of *que* is sensitive to whether the embedding verb is factive or non-factive. Specifically, the *que* in the elliptical clause of a sentence like in (1) above should be largely acceptable under non-factive embedding verbs. Within factive verbs, *que* should display a differential behaviour: under semi-factives, *que* should show a slight degradation in acceptability, and under true factives, *que*

[1] I expect that the results and analysis presented in this chapter generalize to embedding predicates generally, whatever their syntactic category. It would thus be more appropriate to speak of "predicates" here than of "verbs" specifically. However, since my experimental materials on contain verbs as the embedding predicates, I will continue to speak of "verbs". Whenever the topic is more general in this chapter, I will (try to) use the term "predicate". This will be most relevant in Section 3.1.

should show a stronger degradation. Acceptability ratings should nevertheless still be within an acceptable range for both verb types. The reason for these detailed predictions stems from Bîlbîie & de la Fuente's (2019) experimental results on locally embedded gapping, which show this three-way distinction between the verb types. The Typological Conjecture on the relation of globally and locally embedded gapping predicts that Bîlbîie & de la Fuente's (2019) three-way distinction should also surface in globally embedded gapping. I explain the reasoning behind the Typological Conjecture and Hypothesis 1 in detail in Chapter 2 (Sections 2.3, 2.5 and 2.6).

Fernández-Sánchez (2021) – the only contribution in the literature that discusses globally embedded gapping in Spanish at all (as far as I am aware) – specifically addresses the impact of factivity on the acceptability of *que*. He provides two examples, one with non-factive *supongo* 'suppose', see (3), and one with factive *desagrada* 'displeases', see (4). In both examples, *que* is marked as ungrammatical, which leads Fernández-Sánchez (2021) to the conclusion that factivity plays no role for *que*. As I mentioned in in Chapter 2.3, and as I will demonstrate at length in this chapter, these judgments are not accurate, both for the structure in general (i.e. *que* is in fact acceptable), and for the different verb types (i.e. there are differences between these types). I will therefore ignore Fernández-Sánchez's (2021) grammaticality judgments.

(3) Supongo que María traerá las bebidas y (***que**) Pedro ___ el postre.
suppose that María will.bring the drinks and that Pedro the dessert
'I suppose that Mary will bring the drinks and that Pedro will bring the dessert.' (Fernández-Sánchez 2021: 122 f.)

(4) Me desagrada que Pedro me haya servido la sopa fría y (***que**) su mujer ___ el helado derretido.
me displeases that Pedro me has served the soup cold and that his wife the ice-cream melted
'It displeases me that Pedro has served me the soup cold and that his wife has served me the ice-cream melted.' (Fernández-Sánchez 2021: 121, 123)

Let's turn to Hypothesis 2, the difference between gapping and stripping. Stripping, a type of ellipsis closely related to gapping (see Chapter 2.1), has received more attention in the context of global embedding (Villa-García 2016, 2019). In this ellipsis type, the acceptability of *que* seems to be without question, see (5).[2] Stripping therefore may provide a baseline for the gapping data, which is why I include stripping in this acceptability study. As far as I know, the literature does not discuss the influence of the embedding verb on (*que* in) globally embedded stripping.

(5) María dice que a Pedro lo traen, pero **que** a Joaquín no ___.
Mary says that DOM Peter CL bring but that DOM Joaquín not
'Mary says that they will bring Peter, but not Joaquín.' (Villa-García 2016: 141)

Lastly, it is crucial to investigate whether any patterns of *que* I observe in globally embedded gapping are indeed specific to this elliptic structure. Alternatively, it is possible that the *que* patterns surface in

[2]The stripping data discussed here holds specifically for stripping with a polarity particle (like *no* 'not' in 5). Villa-García (2016, 2019) does not discuss other types of stripping. Some of the theoretic literature (Depiante 2004, Villa-García 2016, 2019) treat this type of ellipsis as different from stripping. In Bonke & Repp (2022), we follow this assumption and use Villa-García's nomenclature, calling structures like in (5) "TP-Ellipsis". In this dissertation, I will instead use the term "stripping" to avoid the unnecessary introduction of another technical term.

global embedding generally, regardless of whether the second conjunct is elliptical or not. This is why I included Hypothesis 3 on the difference between elliptical and full clauses in this study.[3]

In total, I conducted the seven acceptability judgment experiments given in the overview in (6). All of these experiments were concerned with Hypothesis 1, the type of the embedding verb. In addition to Factor 1, Experiments 1 and 2 also investigated Hypothesis 2, the potential difference between gapping and stripping. The contrast between Experiment 3, which contained gapping, and Experiment 4, which contained the corresponding full clauses, investigated the issue of ellipsis-specificity, Hypothesis 3.

(6) **Overview over the Experiments**

1. Gapping vs. stripping under indicative non-factives
2. Gapping vs. stripping under subjunctive true factives
3. Indicative non-factives vs. subjunctive true factives in gapping
4. Full clause equivalents of Experiment 3
5. Indicative semi-factives vs. subjunctive non-factives in gapping I
6. Indicative semi-factives vs. subjunctive non-factives in gapping II
7. All four verb types in gapping

Experiments 1 and 2 tested gapping and stripping under non-factives and true factives, respectively. The non-factives in Experiment 1 were verbs of saying and thinking, which select for indicative mood on the embedded verb. This contrasted with the true factives of Experiment 2. These were emotive predicates, which select for subjunctive mood on the embedded verb. The results indicated a differential sensitivity of gapping vs. stripping with respect to the presence of *que* in structures with factive verbs (Experiment 2), but not with non-factive verbs (Experiment 1). In Experiment 2, gapping with *que* was less acceptable with than without *que*, whereas for stripping the presence or absence of *que* did not reduce acceptability. In Experiment 1, there was no difference in acceptability between the various structures. To replicate the results for gapping and compare the two verb types directly, I tested them again in a joint experiment (Experiment 3). The results of this experiment once more showed the sensitivity of *que* to the type of embedding verb: under the non-factives, gapping was equally acceptable with vs. without *que*, while under the factives, the variant with *que* showed some degradation in ratings compared to the variant without *que*.

Experiment 4 tested the full clause equivalents to the gapping items in Experiment 3. The results showed equally high acceptability ratings for both verb types, and for both variants with vs. without *que*. The degradation of *que* under factives in gapping from Experiment 3 thus did not replicate for full clauses, indicating that the degrading effect of *que* is particular to gapping. This result thus provides strong evidence in favour of Hypothesis 3.

In order to facilitate comparison to Bîlbîie & de la Fuente's (2019) data on locally embedded gapping, I tested the third verb type in their data set in Experiments 5 and 6. These were semi-factives, i.e. verbs of knowledge and perception, which take the indicative. I chose to contrast the semi-factives once more

[3] My study on Russian in Bonke (forthcoming) shows why the inclusion of non-elliptical sentences is important: there are subtle, interesting constraints on the complementizer in globally embedded gapping and stripping. However, the same constraints surface in otherwise structurally equivalent non-elliptical sentences, which shows that these constraints are not specific to ellipsis. See Chapter 2.4 for details.

with non-factives. Since Experiments 1 and 3 sufficiently illuminated the pattern of indicative-taking non-factives, I included another set of non-factives in Experiments 5 and 6 instead. These were a somewhat heterogeneous group, including verbs with implicit negation and desideratives. This group of verbs also covered the last possible combination of factivity and mood selection: subjunctive-taking non-factives. The subjunctive-taking non-factives deserve special attention because Bîlbîie & de la Fuente (2019) did not test them. The two verb types in Experiments 5 and 6 showed more nuanced and less reliable patterns w.r.t. the *que*-effect than the indicative-taking non-factives and true factives of Experiments 1 to 3. I will discuss these patterns in detail below. To clear up the resulting uncertainties regarding the semi-factives and subjunctive-taking non-factives, as well as to generate more data for the other two types, Experiment 7 tested all four verb types again.

The combined results of all experiments suggest that the predictions associated with Hypothesis 1 are borne out: (indicative-taking) non-factives show no effect of *que*, true factives show a larger effect, and semi-factives range in between. However, against Hypothesis 1, subjunctive-taking non-factives also show a *que*-effect and thus pattern more closely with the two factive types. While the type of verb thus indeed seems to be a predictor for the acceptability of *que*, factivity is unlikely the determining property. In my analysis, I discuss possible alternatives.

The remainder of this chapter is structured as follows: in Section 3.1, I scrutinize the embedding verbs in my experiments more closely in regard to their two crucial properties: factivity and mood selection. After that, I turn to the description of the materials and design of the experiments (Section 3.2). I present the results of the experiments in Section 3.3. Section 3.4 discusses these results, and Section 3.5 sketches a possible structural account for the behaviour of *que*. Section 3.6 concludes.

3.1 Selection of the clause-embedding verbs

As shown in the overview of the experiments in (6) above, the experiments contained four different verb types. These types are summarized again in the list in (7). They include the three verb types in Bîlbîie & de la Fuente (2019) (indicative-taking non-factives, semi-factives and true factives), plus an additional type, subjunctive-taking non-factives.

(7) **Verb types in the experiments**

1. Indicative non-factives (verbs of saying and thinking)
2. Indicative (semi-)factives (knowledge and perception verbs)
3. Subjunctive (true) factives (emotive verbs)
4. Subjunctive non-factives (desideratives and verbs with implicit negation)

As mentioned in the introduction to this chapter, the reason for including this additional type of embedding verb was to balance the crossing of factivity and mood selection. To explain this in more detail, the two factive sub-types of Bîlbîie & de la Fuente's (2019) study select different moods: the semi-factives select the indicative and the true factives select the subjunctive. The verbs of saying and thinking that represent the non-factives in their study do not show this variable mood selection; they all select the indicative. Thus, the last possible combination of factivity and mood selection, i.e. non-factives that select

the subjunctive, are not present in their study. However, since Spanish has such verbs (the aforementioned desideratives like *desear* 'wish' or *esperar* 'hope/expect', and verbs with implicit negation like *dudar* 'doubt', *temer* 'fear/worry'[4]), including this type seemed like a natural choice.

Table 3.1 lists all the embedding verbs in the experiments in the four verb types. This set of verbs facilitates a comparison with Bîlbîie & de la Fuente's (2019) study, which is the primary goal. But it is also superior to the set of verbs in Bîlbîie & de la Fuente (2019) in some respects: first, as already discussed, it contains the additional verb type that balances factivity and mood selection. Second, each type contains twelve verbs, which is more than Bîlbîie & de la Fuente (2019) test (between three and eight verbs per type, see Chapter 2.5.2). Third, I treat each of the verb types on par, which means that each type contains the same number of verbs. Recall that due to their study design, Bîlbîie & de la Fuente (2019) used a different number of verbs in each type. They did so because the first step in their statistical analysis consisted of a comparison of non-factives to factives generally, with the differences between the factive sub-types being analyzed in a second step. In Chapter 2.5.3, I argue that this procedure is not ideal to carve out differences between the sub-types. My design avoids this problem by treating semi-factives and true factives on the same level with each other and with the two non-factive types. In addition to the balanced number of verbs in each type, the equal treatment of the verb types also shows the experimental design (see Section 3.2 below).

Table 3.1: Verbs of the experiments by their types (in the third person singular present indicative, like in the experimental items).

Non-Factives (with indicative)		**True Factives** (with subjunctive)		**Non-Factives** (with subjunctive)		**Semi-Factives** (with indicative)	
Spanish	*English*	*Spanish*	*English*	*Spanish*	*English*	*Spanish*	*English*
afirma	claims	admira	admires	demanda	demands	acepta	accepts
calcula	estimates	alaba	praises	desea	wishes	admite	admits
considera	considers	apoya	supports	desmiente	denies	averigua	finds out
cree	thinks	aprueba	approves	duda	doubts	conoce	knows
cuenta	tells	consiente	tolerates	espera	expects	descubre	discovers
dice	says	critica	criticizes	niega	denies	ignora	ignores
imagina	imagines	defiende	defends	prohíbe	forbids	nota	notes
piensa	thinks	desaprueba	disapproves	propone	proposes	observa	observes
presume	presumes	detesta	detests	quiere	wants	oculta	hides
relata	relates	lamenta	regrets	rechaza	rejects	oye	hears
sospecha	suspects	odia	hates	recomienda	recommends	sabe	knows
supone	supposes	respeta	respects	teme	fears	ve	sees

A number of syntactic factors restricted the choice of the verbs. First of all, no embedding predicates of other categories (e.g. adjectives like *seguro* 'certain' or *posible* 'possible') entered the sample. All verbs had nominative subjects and took object (rather than subject) clauses. This restriction meant that some of Bîlbîie & de la Fuente's (2019) verbs could not be included in this study, e.g. *gustar* 'like' (literally

[4]The reason for classing *temer* as a negation verb is not immediately obvious from its English equivalents 'fear' and 'worry'. However, the Latin origin of this verb *timeō*, indeed seemed to carry such a negative component to its meaning and selected for the negative complementizer *ne* instead of unmarked *ut* (Gildersleeve & Lodge 1905: 237, §550). For this reason I will group *temer* with the negation verbs within the subjunctive-taking non-factives. (Unlike *temer*, the Latin distinction between negative and non-negative complementizers did not survive in Spanish.)

'be pleased') *sorprenderse* 'be surprised' and *horrorizarse* 'be terrified'. Furthermore, the sample only contained non-reflexive embedding verbs, and verbs that selected for a bare *que*-clause. That is, none of the verbs were such that they came with an obligatory reflexive pronoun or selected for a clause headed by a combination of *que* and an additional preposition. An example of a verb that has both these properties is *quejarse* 'complain', which as shown in (8) selects a complement clause headed by *de* 'of' preceding *que*, and comes with a reflexive pronoun, in this case third person *se*.

(8) María se quejaba de que tenía que hacerlo todo ella.
María CL-3SG complained of that had that do-CL all she
'María complained that she had to do it all herself.' (García Núñez 2021: 1561)

The selection of verbs according to these structural criteria was relatively straightforward. The dictionaries and native speakers I consulted were clear about whether verbs were reflexive or not and whether they embedded prepositional or bare complement clauses.[5] This meant I could make quick and easy decisions on whether each verb in question was a suitable candidate for the experiments. But the verbs also had two other properties which provided more difficulty: factivity and mood selection. Of course, these two properties are decisive for the partition of the verbs into their types. Therefore they deserve special attention. I will discuss them in turn in the following two subsections.

The discussion of factivity (Section 3.1.1) takes a general (i.e. non language-specific) approach and will thus revolve around the literature on English. I assume that the facts in English transfer to Spanish without problems. The discussion will center around the difficulty in obtaining a definite answer to the question whether any given predicate is factive or not, which is due to the variability in interpretation many verbs exhibit. This difficulty is augmented by a similar variability in mood selection, which interacts with factivity in non-trivial ways. I discuss these interactions in Section 3.1.2. These difficulties raise the concern as to how I can be sure that I correctly categorized the verbs into non-factive and factive groups. To anticipate the most relevant reply to this concern, I am – at this point – not worried about the potential inaccuracies in my categorization, since an in-depth empirical investigation of the interpretation of the verbs will follow in Chapter 4. Any inaccuracies or mistakes caused by my categorization will surface there and receive an appropriate discussion.

3.1.1 Factivity

This section outlines my assumptions about factivity (and some related issues) as far as they are relevant to my experiments. I will therefore not attempt to provide an exhaustive overview on the semantic/pragmatic literature on factivity. The central points to the discussion will revolve around two discourse components, the common ground (Stalnaker 1978, 2002) and the Table (Farkas & Bruce 2009). The common ground is the set of propositions whose truth the speaker takes to be uncontroversial in the discourse situation and accepted by all discourse participants (Stalnaker 2002: 151). Such shared propositions in the common ground are called presuppositions (ibid). Propositions may enter the common ground via the second discourse module, Farkas & Bruce's (2009) Table. This module hosts propositions that are currently

[5] I must, however, acknowledge the ample variation between bare *que* clauses and *de que* clauses under the same embedding verb – a phenomenon so common in Spanish that it has its own technical term (*dequeísmo*). See for example Paredes García (2023) for a recent paper on this topic. Since the *dequeísmo* only tangentially affects the verbs in the experiments, I will ignore it for the rest of this dissertation.

under discussion.[6] One way a proposition may enter the Table is via the speech act of assertion, which is the proposal to update the common ground with the propositional content contained in the assertion. If all discourse participants accept the proposition on the Table, it enters the common ground.[7]

This means that the most natural way to introduce new information to the discourse and invite interlocutors to accept this information as shared knowledge (= part of the common ground) is via assertion. Since this subsection is largely concerned with the relation between embedding predicates and their complement clauses, one question concerning the different (types of) embedding verbs in my experiments will revolve around whether it is a possible discourse move for interlocutors to assert the content of the complement clause to these (types of) embedding verbs. To describe this property, I will sometimes use a shorthand formulation such as *the verb asserts (some proposition)*, or speak of a verb *having an assertive reading/interpretation*.

I assume that the factive verbs in my experiments both presuppose and entail the propositional content of their complement clause. In doing so, I follow much of the literature on factivity (e.g. Gazdar 1979, Chierchia & McConnell-Ginet 1996, van der Sandt 1992, Schlenker 2010, Anand & Hacquard 2014, Spector & Egré 2015).[8] I follow Degen & Tonhauser (2022) in abbreviating the propositional content of the complement clause (to a contextually given predicate or set of predicates) by 'CC'. By presupposing the CC I mean that the utterer of a sentence S containing a complement clause takes the truth of the CC for granted and uncontroversial in discourse. This means that a condition on the utterance of S is that the truth of the CC be established in the common ground, or be easily accommodated into the common ground. To illustrate, consider the example in (9). When someone utters the sentence in (9), which contains the factive verb *regrets*, they presuppose that the proposition *it is raining* (= the CC of *regret*) is true.

(9) I regret that it is raining. (Kiparsky & Kiparsky 1970: 147)

As a consequence of being part of the common ground, the CC of factives is backgrounded in discourse and not at-issue (see e.g. Simons et al. 2010, 2017, Abrusán 2011, Anand & Hacquard 2014, Tonhauser et al. 2018, among others). These two properties, backgrounding and non at-issueness, affect what contributions in discourse are appropriate with factive and non-factive predicates. Factives cannot introduce their CC as new information to the discourse, i.e. they cannot put their CC on the Table. This entails that factives cannot assert their CC. To illustrate, consider the question-answer pair in (10) and (11). The answer to the question (10) is contained in the CC of non-factive *think* in (11a) and factive *regret* in (11b). The answer with *think* is an appropriate utterance, because the CC is not presupposed and therefore is not given in the common ground. Consequently, the CC contains information that is new and at-issue within the discourse. The utterance in (11a) is thus an assertion of the CC. The opposite is the case for the answer with *regret*, since this verb requires its CC to be present in the common ground upon utterance.

[6] One way to approach propositions that are on the Table is in a question-based framework, where such a proposition is treated as 'Question under Discussion' (QUD) (Roberts 1996, 2012). I touch upon this approach in my discussion of Bîlbîie & de la Fuente's (2019) analysis. It will play no role here.

[7] Both of these modules, common ground and the Table, have received formal implementations in set-theoretic terms. For the discussion at hand, such a formalism is not required.

[8] The earliest treatments of factivity (Kiparsky & Kiparsky 1970, Karttunen 1971) make do with presupposition alone, without mentioning entailment. See Degen & Tonhauser (2022) for a discussion.

The CC thus cannot be presented as new, at-issue information (= asserted), and therefore cannot serve as the answer to a question.[9]

(10) How's the weather in Cologne?

(11) a. I think that it is raining in Cologne.
b. # I regret that it is raining in Cologne.

In a similar vein, the non at-issueness of the CC of factives means that it cannot be straightforwardly targeted for rejection or acceptance in discourse. To agree or disagree with the CC, it takes more effort on the part of the interlocutor if the predicate is factive than if it is non-factive. This difference shows between non-factive *think* in (12) and factive *hate* in (13). The non-factive verb easily allows rejection (12a) or acceptance (12b) of the CC by the interlocutor. The factive verb does not allow rejection (13a) or acceptance (13b) of the CC as easily. To target the CC under *hate* for rejection, the interlocutor must put a more substantial effort into their contribution to the dialogue, see (13c). This utterance exemplifies the so-called 'Hey wait a minute' test (Shanon 1976), which is a classical diagnostic for presuppositions.

(12) A: I think it's raining in Cologne.
a. B: No, that's not true, it isn't raining in Cologne.
b. B: Yes, that's true, it is raining in Cologne as always.

(13) A: I hate that it's raining in Cologne.
a. B: #No, that's not true, it isn't raining in Cologne.
b. B: #Yes, that's true, it is raining in Cologne as always.
c. B: Hold up, wait a minute, something isn't right. I don't think it is actually raining in Cologne.

Presupposition of the CC, as displayed by these properties, will be the central aspect of factive predicates in the discussion of my experimental data. The second property associated with factive predicates is the entailment of the CC. Entailment means that the truth of the CC is an indefeasible meaning component of the sentence that contains it (Gazdar 1979, Chierchia & McConnell-Ginet 1996, see also Degen & Tonhauser 2022). Thus, in (11b) above, the CC of factive *regret* is entailed as well as presupposed. As a consequence, a continuation of this sentence like in (15) that calls the truth of CC into question, leads to a contradiction. With non-factive *think*, the CC is not entailed, which is why such a continuation is possible in (15).

(14) I think that it is raining in Cologne, but it might not be raining in Cologne.

(15) I regret that it is raining in Cologne, #but it might not be raining in Cologne.

A core property of factive presuppositions (as of any presupposition) is that they "survive" in entailment cancelling environments. A sentence provides such an environment if it contains operations that manipulate the truth conditions of an equivalent simple, affirmative sentence such that the entailments within

[9] While this pattern generalizes for the two factive groups in my experiments, there is more to be said about the non-factives. Not all non-factives can assert their CC, some of which pattern like factives in this regard. The same is true for the point about rejection or acceptance of CCs in discourse I make in the next paragraph. I will return to the heterogeneous behaviour of non-factives below, and for the time being only consider those non-factives that pattern as described.

that simple sentence disappear. The sentences in (16) show what this means: the truth conditions of the simple assertive sentence in (16a) are fundamentally different from its negated version in (16b), the polar question in (16c), the conditional clause in (16d), and the version with a modal expression in (16e). Entailments of the simple sentence (16a), e.g. *Someone regrets something*, are not present in the other sentences (hence entailment cancelling environment). However, like the simple sentence in (16a), the sentences in (16b) to (16e) still presuppose *it is raining*, i.e. the CC of the factive embedding verb *regret*. This "survival" of the factive presupposition is referred to as "projection": the presupposition projects out of the entailment cancelling environments. Projection is commonly tested by means of a group of sentences with different entailment cancelling environments like in (16b) to (16e). This test is called the "family-of-sentences test" (Chierchia & McConnell-Ginet 1996).

(16) a. Simple sentence: I regret that it is raining. (same as 9 above)
b. Negation: I don't regret that it is raining.
c. Polar question: Does anybody regret that it is raining?
d. Conditional clause: If anybody regrets that it is raining (I will not be surprised).
e. Modal expression: Maybe somebody regrets that it is raining.

The projection of out of these entailment cancelling environments differentiates factive predicates from another set of predicates, which like factives entail, but unlike factives do not presuppose their CC. These predicates are called veridical predicates (Giannakidou 1998, Egré 2008). All factive predicates are thus also veridical, but not vice-versa (Giannakidou 1998: 150). An example for a veridical predicate is *be right* in (17) and (18). The unmodified assertive clause in (17) entails the CC, i.e. that Mary is the murderer, and thus patterns like a factive in this regard. However, the CC is not presupposed under this predicate since it does not project out of entailment cancelling environments. The lack of projection is shown in the negated version of (17) in (18), which does not entail the truth of the CC. The predicate *be right* is thus veridical and not factive.

(17) John is right that Mary is the murderer. (Anand & Hacquard 2014: 72)

(18) John isn't right that Mary is the murderer. (Anand & Hacquard 2014: 72)

I assume the two core properties of factives, presupposition and entailment of the CC, for all verbs of the two factive sub-types in my experiments. For the two non-factive types, I assume that these two properties do not hold. This is not to say that non-factives might not trigger other presuppositions that concern their CC. In fact, especially the subjunctive-taking non-factives (desideratives and verbs with implicit negation) have been argued to be presupposition triggers. The negation verbs (e.g. *desmiente*, *niega*, both 'denies', or *rechaza* 'rejects') are only felicitous in a context in which the CC is already under discussion. These verbs thus "presuppose the existence of a discourse referent as their complement" (Kastner 2015: 157), which means that the CC has to be on the Table, but not yet part of the common ground. These verbs thus may serve as responses to previous assertions of their CC (which is why these verbs are sometimes called 'response stance' verbs, e.g. Weir 2014: 234). As a consequence of this discourse condition on the use of these verbs, I assume that they cannot themselves assert their CC. Support for this assumption

comes from the fact that the CC of a negation verb like *deny* cannot function as the answer to a question see (19). In this regard, non-factive *deny* thus patterns like factive *regret*, (cp. 11b).

(19) a. How's the weather in Cologne?
b. # I deny that it is raining in Cologne.

The other sub-class within the subjunctive-taking non-factives are desideratives such as *quiere* 'wants' or *propone* 'proposes'. These verbs have been argued to trigger or interact with presuppositions as well (Heim 1992). Meinunger (2017) proposes that these verbs are counter-factive, i.e. that they presuppose the negation of their CC. I will not go into detail on Meinunger's (2017) proposal, since the (presumed) counter-factivity of desideratives is not important for the issue at hand (and also I do not find the proposal convincing). I will, however, assume that like negation verbs, desideratives also cannot be used to assert their CC (whether this is due to a presupposition they trigger or not). Thus, the subjunctive-taking non-factives generally cannot introduce new information to the discussion with their CCs. In this aspect, they pattern with the factives, but unlike the factives, they do not presuppose the truth of their complement. As a result of these assumptions only the indicative-taking non-factives in my experiments are such that their CC can be at-issue and thus asserted.

While the criteria for factivity are relatively clear (presupposition plus projection, and entailment), it is not always trivial to say whether these criteria apply to an individual clause-embedding predicate. This is to say that the classification of a predicate as factive or non-factive may be tricky. Different contributions in the literature on factivity may put the same predicates into different categories or assumed sub-categories. For a general overview of these inconsistencies in classification, I refer to Degen & Tonhauser (2022) and references therein. I will spend the rest of this subsection on illustrating these difficulties of categorization on the basis of two examples: the verb *know*, and the class of emotive predicates. These predicates are especially relevant to the discussion because my experimental materials contain them: the Spanish equivalent of *know*, *sabe*, is one of the semi-factives (and represents this verb type for the purposes of this discussion), and the true factives are emotives.

In philosophy, what it means to 'know' something has long been treated under the standard definition of 'justified true belief' (see e.g. Ichikawa & Steup 2018 for an overview over philosophical analyses of knowledge). With the notion of 'truth' baked into the common definition of knowing, it comes as no surprise that the corresponding clause-embedding verb *know* is often treated as a representative token of a factive predicate in the linguistic literature (e.g. Kiparsky & Kiparsky 1970, Karttunen 1971, Hooper 1975, Chierchia & McConnell-Ginet 1996, Egré 2008, Abrusán 2011, Anand & Hacquard 2014, Meinunger 2017, among many others). At first glance, this categorization is justified, since *know* seems to easily pass the family-of-sentences test. Both the unmodified assertive clause in (20) as well as the sentences with entailment cancelling environments in (21) to (24) presuppose the CC *it rains a lot in Cologne*.

(20) Mary knows that it rains a lot in Cologne.

(21) Mary doesn't know that it rains a lot in Cologne.

(22) Does Mary know that it rains a lot in Cologne?

(23) If Mary knows that it rains a lot in Cologne (she won't visit.)

(24) Maybe Mary knows that it rains a lot in Cologne.

However, the literature features a number of counterexamples in which *know* does not entail and/or does not presuppose its CC. Beaver (2010) provides a list of actual use examples of non-projecting *know*. One of Beaver's (2010) examples is in (25), where the modal expression *perhaps* provides the entailment cancelling environment. The second sentence in this example shows that the speaker in this situation does not take the truth of the CC for granted, which they say explicitly (*I don't believe this, but you never know*). The presupposition of the CC thus does not project from under the modal expression, contrary to what should be happening if *know* were factive. Under the right discourse conditions, then, it seems that a sentence with *know* may receive an interpretation in which the CC is not presupposed, i.e. a non-factive interpretation.

(25) […] It seemed obvious for him to continue trumps, but perhaps he **knows** that diamonds are running and is trying to set up a trump spot for partner. I don't believe this, but you never know. (Beaver 2010: 86, emphasis mine)

Another instance in which *know* may receive a non-factive interpretation is in (26). The paraphrase of this sentence in (27) shows that there is a reading of (26) in which the CC is not presupposed. In contrast to (25) above, the non-factive interpretation seems to not be tied to discourse conditions, but instead to structural properties of the matrix clause (first person subject and negation, Kiparsky & Kiparsky 1970: 148). Again, the availability of such a non-factive interpretation of *know* is unexpected under the assumption that *know* is factive.

(26) I don't know that this isn't our car. (Kiparsky & Kiparsky 1970: 148, fn. 4)

(27) This might be our car.

The literature features a number of approaches to the (apparent) non-factive potential of *know* in sentences such as (25) and (26). I will focus on one such approach, which splits the class of factive predicates into two sub-classes, semi-factives and true factives. This split is present in Bîlbîie & de la Fuente's (2019) study, and I follow it (at least in nomenclature, see below) in my experiments as well. As the name suggests, the split divides factive predicates into two classes, depending on how reliably they display factive inferences. Within this split, true factives (supposedly, see below) show consistent and invariant projection, while semi-factives show variability in their projection behaviour (see Karttunen 1971, Hooper & Thompson 1973, Hooper 1975 and much subsequent literature).

This split was proposed by Karttunen (1971), who also coined the term 'semi-factive'. He argues that the truth of the CC of semi-factive predicates possibly, but not necessarily, projects from under some specific entailment cancelling environments. The "not necessarily" is crucial here since it allows for variability in projection behaviour. His discussion features the verbs *regret, realize* and *discover*, which he claims invariably project a factive presupposition from under negation, see (28).

(28) John didn't {regret / realize / discover} that he hadn't told the truth. (Karttunen 1971: 63)

This projecting behaviour in an entailment cancelling context in principle qualifies all three verbs as factive, and raises the expectation that they show identical projection patterns in all other entailment

cancelling environments as well. However, Karttunen (1971) argues that this expectation is not met for all these verbs. Only *regret* receives factive readings in questions like (29) or conditionals like (30), while the other two verbs, *realize* and *discover*, show more nuanced patterns. Specifically, he argues that in questions like (29), these two verbs may receive an interpretation that is either factive or non-factive, and in conditional clauses like (30), the interpretation of these two verbs is categorically non-factive. The consistent projection with *regret* makes this verb a true factive,[10] while the variability in projection with *realize* and *discover* makes them semi-factives. Karttunen does not discuss the verb *know* explicitly, but he mentions that the group of semi-factives encompasses "*discover, realize, find out, see, notice* and other similar verbs" (Karttunen 1971: 65), which plausibly extends to *know* as well. In any case, the variable projection behaviour of *know* discussed above qualifies it for the class of semi-factives. Based on Karttunen's (1971) list of individual semi-factive verbs, I therefore assume (with Bîlbîie & de la Fuente 2019) that semi-factives are verbs of knowledge and perception.[11]

(29) Did you {regret / realize / discover} that you had not told the truth? (Karttunen 1971: 63)

(30) If I {regret / realize / discover} later that I hadn't told the truth, I will confess it to everyone. (Karttunen 1971: 64)

On the other side of the split within factive predicates are the true factives. I follow Hooper (1975), Beaver (2010) and Bîlbîie & de la Fuente (2019) among others in assuming that these are emotive predicates.[12] One such emotive (true) factive is *regret*, which features prominently in the literature as a representative token of this type of predicate (see e.g. the example from Karttunen 1971 above). From Karttunen's (1971) description of the projection patterns of *regret*, it is to be expected that this verb, and by extension all verbs of its type, display invariant presupposition projection in all discourse contexts and in all entailment cancelling environments. However, this expectation is not borne out in all contexts.

One such context involves false beliefs. An example of such a context is in (31), which features the emotive factive verb *regret*. In this example, the CC (*Oedipus killed the stranger*) is explicitly described as false in the initial gerund clause. (The CC is thus not entailed either.) This example shows that, like for *know* and other semi-factives, the presupposition with supposedly true factives may be suspended in the appropriate discourse contexts.

(31) Falsely believing that he had inflicted a fatal wound, Oedipus regretted killing the stranger on the road to Thebes. (Klein 1975, as cited in Degen & Tonhauser 2022: 557)

This suspension of the factive presupposition in false belief contexts also applies to semi-factives (Abrusán 2011). The second sentence in (32) contains our token semi-factive *know* in such a context. Like in example (31) above, the truth of the CC is not presupposed or entailed because the context strongly suggests that the CC is false. A similar situation is present in (33), where the falsity of the belief is not explicitly

[10] Karttunen (1971) only uses the terms 'semi-factive' and 'factive' to refer to the two different factive sub-types. The term 'true factive' is from Hooper & Thompson (1973) and Hooper (1975). I follow Bîlbîie & de la Fuente (2019) in using this term.

[11] For an account that suggests that semi-factives are actually not factive at all and carry no presupposition whatsoever (for Spanish, specifically), see Klein (1977).

[12] Beaver (2010) acknowledges the split of factive between emotives and knowledge/perception verbs, but does not associate these verbs with the terms 'semi-factive' or 'true factive'. Emotives themselves have been treated as factives at least since Kiparsky & Kiparsky (1970) and Karttunen (1971); I focus here on their status as a well-defined sub-type within factives.

mentioned, but transported via a marked intonation on the verb *knew*, which indicates its "scare quote use" (Egré 2008: 103). Again, the CC is neither entailed nor presupposed in this example.

(32) John suffers from paranoia. He falsely believes that the police is spying on him and what is more he knows they are listening to his phone calls. (Abrusán 2011: 514)

(33) Before Copernicus, people "knew" that the sun was revolving around the earth. (Egré 2008: 103, I put quotation marks around *knew* to indicate scare quote intonation.)

It is not entirely clear from Egré's (2008) exposition what the prosodic exponents of this "scare quote" intonation would be. However, whatever this type of special intonation entails, it does not seem to be the only general point of connection between prosody and (the suspension of) factive inferences: as put forward by Beaver (2010) and Simons et al. (2017), and confirmed experimentally by Tonhauser (2016) and Djärv & Bacovcin (2017), if a constituent in the complement clause to a factive predicate is prosodically prominent (e.g. by accentuation), the CC is less likely to project than if the complement clause is deaccented. For an example, see the dialogue in (34). In this dialogue, B's utterance contains the factive embedding predicate *find out* under an entailment cancelling operator (negation). Generally, the CC of this predicate is thus expected to project. However, the complement clause contains the prosodically prominent constituent *HARRY* (capital letters indicate prosodic prominence). As a result, the CC in B's statement is less likely to be understood to be presupposed and/or entailed. This effect of presupposition suspension is not an artifact of the prosodic properties of the complement clause as such, but instead of the information structural properties that prosody marks. For a more in-depth discussion of the interaction of information structure, prosody and factivity, see Simons et al. (2017).

(34) A: James just found out that Harry's having a graduation party, and I just can't understand why he's so upset about it.
B: He didn't find out that HARRY's having a graduation party (he found out that HARRIET is having a graduation party). (adapted from Simons et al. 2017: 190)

Returning to false belief contexts,[13] these contexts show that emotive predicates must be included in the group of predicates that show variable, rather than categorical, factivity – a conclusion that leaves no categorically factive predicates. Factive inferences seem to be suspendable, in one way or another, for all predicates. For the emotives specifically, one possible conclusion is that they are actually not factive. This position is held by a part of the literature (Klein 1975, Heim 1992, Egré 2008). Other contributions maintain that emotives are factives (Gazdar 1979, Abrusán 2011, Anand & Hacquard 2014), and account for their behaviour in false belief contexts in other ways. The general idea is that in these specific contexts where the falsity of the embedded proposition is made explicit (like in 32) or conveyed via other means (e.g. intonation in 33), the factive presupposition is attributed to the matrix subject and not to the speaker (Karttunen 2016: 712). The resulting interpretation of sentences like (32) and (33) is that of a "kind of free indirect discourse" (Anand & Hacquard 2014: 85) or "free indirect speech" (Abrusán 2011: 514), which I interpret as a report of the subject's beliefs and the subject's attitudes towards those beliefs.

[13] Note that the dialogue in (34) above also involves a false belief, i.e. A's (apparent) belief in the CC. It is unclear whether such a false belief context is a necessary prerequisite for prosodic prominence to contribute to a suspension of factive inferences in the manner described above. A more detailed discussion of the role of false belief in the prosody-information-factivity interaction is well beyond the scope of this section.

Despite these observations, I will continue to assume that semi-factives and true factives presuppose their CCs. I thus essentially disregard the impact of the data presented in this section on the verb types, i.e. the variable projection behaviour in different entailment cancelling environments (for semi-factives), and the false belief contexts (for both factive types). I do so for two reasons: first, my experiments contain unmodified declarative sentences only, i.e. sentences without entailment cancelling environments,[14] and without false belief contexts. The patterns in these specific environments do thus not directly affect my experimental materials. Second, Degen & Tonhauser's (2022) rating study on (English) embedding predicates finds no categorical difference between presumed factive sub-types as far as their presupposition projection is concerned. In Chapter 4, I conduct a rating study on Spanish, which is closely modelled after Degen & Tonhauser (2022) (which is also where I will go into detail on their study). The results of the study in that chapter paint a similar picture to Degen & Tonhauser's (2022): there seems to be no categorical difference between semi-factives and true factives w.r.t. presupposition projection. Even though I thus disregard much of what I have discussed until this point, I chose to include the discussion in this subsection in order to show the variable behaviour that all seemingly factive predicates display in some environment or other, and the resulting difficulties in the categorization of embedding predicates as factive or non-factive.

My assumption that semi-factives and true factives are both equally factive (i.e. both presuppose and entail the content of their complement) raises the question as to how they are different. A formal difference is of course mood-selection, which I already briefly touched upon in Chapter 2.5, and to which I will return in the following subsection. What is more relevant now is whether the two types differ as to their meaning, especially as concerns the CC. Recall from the discussion of Bîlbîie & de la Fuente's (2019) experimental results (Chapter 2.5.3) that they make the assumption that the two types differ in assertiveness. I addressed the potential of the different verb types to assert their CC above already, but I will repeat my points here to make the differences between Bîlbîie & de la Fuente's (2019) and my assumptions explicit.

Bîlbîie & de la Fuente (2019) assume, with Hooper (1975) and Farkas (2003), that semi-factives carry a feature [+assertive] and true factives carry a feature [–assertive]. By [+assertive] I understand that a verb encodes at-issue content in its complement clause,[15] which is open for acceptance or rejection into the common ground, i.e. it is on the Table. Verbs that are [–assertive] encode non-at-issue content, which in discourse is not available for acceptance or rejection into the common ground. The idea that factive verbs may be [+assertive] in this sense clashes with the assumption I made above that factive presupposition goes with backgrounding of the complement proposition, which therefore is not at-issue. For this reason I reject the idea that any factives are categorically factive and assertive at the same time.

It still seems possible that under certain circumstances, the complement clause may receive an assertive interpretation, if at the same time, the factive presupposition and entailment are suspended (Terrell 1976, Klein 1977, Lleó 1979).[16] However, I do not assume that this possible shift in interpretation is characteristic

[14]The absence of entailment cancelling environments in the experimental items raises the question why factivity (rather than veridicality) should play a role for these items in the first place. I acknowledge this issue, and will return to it in Chapter 5. For now, I keep following Bîlbîie & de la Fuente (2019), who do not address this question either.

[15]This assumption conflates the distinct notions of at-issueness and assertion. I will proceed with this assumption under the condition that it applies only to declarative complement clauses. I will return to at-issueness (and assertion) in embedded questions in my structural analysis of the experimental results in Section 3.5.

[16]This is what Hooper (1975) may have had in mind when categorizing semi-factives as both assertive and presuppositional.

of semi-factives but it instead may apply across the board for factives and in the right circumstances. My assumption on the default interpretation on both factive sub-types is that they are non-assertive, which means that their CC cannot introduce new or at-issue content to the discourse.

In essence, then, the only difference I assume between semi-factives and true factives is in mood selection. As far as the basic meaning components relevant to this discussion (= factivity and assertiveness) are concerned, these two verb types are identical. I will still use the terms 'semi-factive' and 'true factive' as descriptive labels for these two verb types, for ease of exposition, and to facilitate comparison to Bîlbîie & de la Fuente's (2019) study.

Summing up my assumptions about the two types of factive verbs in my experiments: I assume their CCs are presupposed and entailed, and project from under entailment cancelling environments (as is standard for presuppositions). I further assume that the CCs are not at issue and and (therefore) not asserted. However, I assume that these properties are not set in stone, but potentially variable. The predicates may (for example) receive free indirect speech interpretation, depending on the discourse environment. (This shift in interpretation will play a role later.) As for the non-factives, I assume that they do not presuppose and entail their CCs. However, I assume that only the indicative-taking non-factives allow for an assertive or at-issue reading (= only these verbs can put their CC on the Table). The subjunctive-taking non-factives generally lack this ability, however as for the factives, I leave open the possibility they may show some variability in these properties (but I did not discuss this last aspect explicitly).

As the individual verbs are concerned, I checked for their core properties (presupposition and entailment of the CC) by means of my own introspective judgments. I would like to make clear at this point that I am aware that this procedure will likely have lead to inaccuracies for at least some of the verbs. The reason for these inaccuracies are the general difficulties I outlined in this section that surface when trying to categorically decide whether or not some predicate is factive. These difficulties were augmented by the fact that I went on the German and English translation equivalents of the Spanish verbs, which was necessary due to the fact that my practical knowledge of this language is more than limited. To remedy this significant shortcoming, I will devote Chapter 4 to an in-depth experimental examination of the factivity of these embedding verbs. I am thus – at this point – unconcerned of the inevitable problems of the categorization procedure.

3.1.2 Mood selection

The second dimension along which the embedding verbs in my experimental materials vary is mood selection: one half of the verbs select indicative clauses and the other half select subjunctive clauses. Mood in Spanish is an inflectional category of the finite verb. The two different moods are illustrated in (35), which has the indicative form of the embedded verb (*llueve* 'it rains'), and in (36), which has the subjunctive form (*llueva* 'it rains'). This difference in mood is triggered by the two embedding verbs, *creo* 'think' and *lamento* 'regret' in (35) and (36), respectively, which select for the different moods.

It is not entirely clear whether Hooper (1975) actually assumes that CCs of semi-factives are *simultaneously* asserted and presupposed, or whether they may receive either assertive or presupposed readings, depending on context (Hooper 1975: 117). About the apparent clash of these two properties, she writes "I must say that either presupposition needs to be redefined to exclude this weaker type of presupposition [= the presupposition of semi-factives], or we must admit that weakly presupposed propositions may also be asserted" (Hooper 1975: 117). I see no such redefinition of factivity in the literature that provides a satisfactory solution to this clash.

(35) Creo que **llueve** todo el tiempo en Colonia.
think-1SG that rain-IND all the time in Cologne
'I think that it rains all the time in Cologne.'

(36) Lamento que **llueva** todo el tiempo en Colonia.
regret-1SG that rain-SUBJ all the time in Cologne
'I regret that it rains all the time in Cologne.'

In the experimental materials, I followed the intuitions of my informants (native speakers of European Spanish) when I decided what verb should go with what mood. My informants expressed strong preferences for all verbs in this regard, which makes me confident that the experimental items are grammatical and acceptable (as far as mood is concerned). There are nevertheless some issues concerning mood selection that need discussing.

Much like factivity, mood (in general, as well as in Spanish specifically) has received attention from many different angles in the linguistic literature. Also like for factivity, I will not attempt to give an exhaustive overview over these different angles. Instead I focus on the possible interactions that mood selection has with factivity. This entails that I limit the discussion to mood in complement clauses and disregard the indicative-subjunctive divide in main, adverbial and relative clauses. One specific aspect I will discuss is whether there is variability in mood selection, i.e. whether the same predicate may select for both indicative and subjunctive complement clauses, and if so, whether this variability affects the (non-)factive interpretation of this predicate. This issue relates directly to the design of my experimental materials: if there is a choice to be made as to whether any given predicate takes the indicative or the subjunctive, this choice might influence the interpretation of this predicate, and in turn its categorization as factive or non-factive.

The idea that mood selection in Spanish (and other Romance languages) generally relates to notions such as presupposition and assertion is widespread in the literature (Bolinger 1968, Terrell & Hooper 1974, Lunn 1989, Mejías-Bikandi 1994, Schlenker 2005, Meinunger 2017, among others). A general tenet in the literature is that the subjunctive marks non-assertive CCs,[17] which applies straightforwardly to the two subjunctive-taking verb types in my experiments: the factives (i.e. emotive, true factives) presuppose the their CC and thus do not assert it. The non-factives (verbs with implicit negation and desideratives) do not presuppose the their CC, but also do not assert it, which possibly has to do with other kinds of presuppositions they trigger. However, the correspondence between mood selection and non-assertion is not absolute. The semi-factives select for the indicative, but like the true factives presuppose their CC and therefore do not assert it.[18] See Section 3.1.1 above for a more detailed discussion on the assumptions I make on the verb types and the presuppositions they trigger.

Thus, non-assertive CCs are distributed across verb types that may select either the subjunctive (the true factives and subjunctive-taking non-factives in my experiments) or the indicative (the semi-factives). This distribution suggests that it is unlikely that the moods themselves carry a meaning that outright determines the interpretation of CCs by signalling assertion, presupposition generally, or factivity specifically. Lexically determined selection seems to play a considerable role in the distribution of mood:

[17] CC = the propositional content of the complement clause to a given predicate or set of predicates. I introduced this abbreviation in Section 3.1.1 above, following the terminology in Degen & Tonhauser (2022).

[18] See Terrell & Hooper (1974: 486) for an opposing view.

whether a given predicate selects for the indicative or the subjunctive seems to be in large parts a fixed property tied to lexical features of these predicates such as emotiveness, desiderativeness and so on. These lexical features seem to have a greater impact on the function of the CCs of these predicates in discourse such as presupposition and assertion than the moods they select.

However, mood selection cannot be entirely relegated to lexical aspects of this kind since one and the same predicate may vary to some degree as to what mood it selects for. For example, I classed the verbs of saying in my experiments as taking the indicative (as well as being non-factive), but these verbs also may, under certain conditions, come with a subjunctive complement. Two widely discussed instances of verbs of saying with subjunctive complements are when they embed a directive or a desideratum (rather than an assertive clause), and the so-called 'polarity subjunctive' (Quer 1998, 2009, 2011). I discuss these two phenomena in turn.

An example of directive embedding under a verb of saying is in (37). The verb is *dile* (the imperative form *di* of *decir* 'say' plus the indirect object clitic *le* 'him'), which takes a complement clause with the subjunctive verb *venga* 'come'. The resulting interpretation is *Tell her that he should come*, i.e. an instruction of the speaker to his addressee (via the main clause imperative) to deliver an order (the embedded subjunctive) to a third person (the referent of the indirect object clitic). When the embedded clause contains an indicative verb such as *viene* 'comes' in (38), (= the indicative counterpart to 37), the clause receives a declarative interpretation (*Tell her that he is coming*). The fact that the subjunctive leads to a directive or desiderative reading explains why predicates that invariably carry such a directive or desiderative meaning select for the subjunctive. This concerns the desideratives in my experiments, i.e. a part of the subjunctive-taking non-factives.

(37) Dile que venga.
say.IMP-CL that come-3SG.SUBJ
'Tell him that he should come.' / 'Tell him to come.'

(38) Dile que viene.
say.IMP-CL that come-3SG.IND
'Tell him that he is coming.'

The second phenomenon in which verbs of saying and thinking take the subjunctive is the so-called 'polarity subjunctive'. This term refers to the fact that verbs of saying and thinking may take a subjunctive complement clause if the matrix clause is negated (Rivero 1971, Lleó 1979, Quer 1998, 2009, 2011). This switch from indicative to subjunctive does not only occur under negation, but also in polar questions (Rivero 1971), certain modal environments (Fábregas 2014), or with quantified matrix subjects (Laskurain–Ibarluzea 2015). I concentrate on the pattern with negation, which is illustrated in (39) for the embedding verb *cree* 'thinks'. The positive version of this sentence (39a) features the indicative verb *ayuda* 'supports' in the complement clause, which is in accordance with the general selection principles of this verb. With negation in the matrix clause, both moods are grammatical in the complement clause: indicative *ayuda* as in (39b), and subjunctive *ayude* as in (39c).

(39) a. **Positive matrix clause with indicative complement clause**
Vladimir cree que la Unión Soviética **ayuda** al régimen de Videla.
Vladimir thinks that the union Soviet supports-IND at-the regime of Videla
'Vladimir thinks that the Soviet Union supports the Videla regime.'

b. **Negative matrix clause with indicative complement clause**
Vladimir **no** cree que la Unión Soviética **ayuda** al régimen de Videla.
Vladimir not thinks that the union Soviet supports-IND at-the regime of Videla
'Vladimir doesn't think that the Soviet Union supports the Videla regime.'

c. **Negative matrix clause with subjunctive complement clause**
Vladimir **no** cree que la Unión Soviética **ayude** al régimen de Videla.
Vladimir not thinks that the union Soviet supports-SUBJ at-the regime of Videla
'Vladimir doesn't think that the Soviet Union supports the Videla regime.'

(Lleó 1979: 166, my glosses, translations, and headlines)

With matrix negation, the subjunctive (39c) seems to be the more natural in the majority of contexts, while the indicative (39b) requires a more specific context and triggers a special interpretation (Lleó 1979). Lleó (1979: 166 f.) points out that while both variants constitute a report on the subject's mental state (i.e. Vladimir's not believing the CC is true in 39b/39c), the version with the indicative additionally entails a "claim to truth" of the CC on part of the speaker. With the indicative, the speaker signals their commitment to the truth of the CC, while with the subjunctive, the speaker may or may not be committed to the truth of the CC. An example of how this additional inference for the indicative becomes visible is in the three sentences in (40). Like in (39), the sentences contain negated matrix *cree* 'think' and complement clauses with the different moods. The crucial aspect in these examples is that they also feature a continuation in which the speaker agrees with the negative belief of the matrix subject *Juan*, thus rejecting the CC. Such a continuation is fine with the subjunctive in (40a), but leads to a contradiction with the "claim to truth" of the CC conveyed by the indicative in (40b).[19] A continuation for the version with the indicative would have to imply that the speaker is certain that the CC is true, thus disagreeing with the belief state of the matrix subject. This is the case in (40c).

(40) a. Juan no cree que María **venga** y yo lo creo también.
John not thinks that Mary comes-SUBJ and I it think too
'John doesn't believe that Mary is coming and I believe the same thing (i.e. that Mary is not coming).'

b. * Juan no cree que María **viene** y yo lo creo también.
John not thinks that Mary comes-IND and I it think too
'John doesn't believe that Mary is coming and I believe the same thing (i.e. that Mary is not coming).'

c. Juan no cree que María **viene** pero yo lo creo.
John not thinks that Mary comes-IND but I it think
'John doesn't believe that Mary is coming but I do believe so (i.e. that Mary is coming).'

(adapted from Rivero 1971: 310 f., I added glosses and the bracketed parts to the translations)

This "claim to truth" aspect to the meaning of the indicative raises an obvious question: how does it relate to the properties that are relevant to my discussion, (factive) presupposition and assertion? This

[19] Rivero (1971) marks her example with the indicative as ungrammatical rather than semantically anomalous. However from her exposition it is clear that is the meaning and not grammar that causes the problem in (40b). Note furthermore that Klein (1977: 80) and Lleó (1979: 12, 26) consider both (40a) and (40b) ungrammatical. They unfortunately do not discuss why this discrepancy in judgments might be (expect for potential dialectal differences). To me, it seems possible that Klein's (1977) and Lleó's (1979) rejection of these examples stem from the lack of clarity how the propositional anaphors in these examples relate to the negation in their antecedent (see the awkward English translations, for which I added the clarifying parts in brackets), and not the actual pattern the examples seek to illustrate. I therefore assume the actual differences between the moods in (40) are genuine.

question has received divergent answers in the literature. Rivero (1971) argues that the indicative triggers a factive presupposition (which is interesting considering that above we saw that the subjunctive is often considered an indicator of presupposition). Klein (1977), making reference to Rivero (1971) says that it cannot be presupposition but must be assertion instead. Lleó (1979), making reference to both Rivero (1971) and Klein (1977), says that it is neither presupposition nor assertion, and instead argues for entailment. I will not go into detail on these three contributions, partially because the definitions of the crucial notions of presupposition and assertion differ between these contributions, and are also partially different to my definitions from Section 3.1.1. However, the impossibility of negating the CC in a continuation like in (40b) above represents what I introduced as a test for entailment in Section 3.1.1. The diagnostic for (factive) presupposition on the other hand was projection out of entailment cancelling environments (the 'family-of-sentences test'). The polarity subjunctive and its alternation with the indicative is itself tied to at least some of these entailment cancelling environments (negation, question formation, modal expressions), which effectively makes it impossible to conduct the 'family-of-sentences test'. This circumstance in turn makes it impossible to tell for sure whether what we observe in (39) is connected to presupposition or entailment.

Summing up so far, even though I class the verbs of saying and thinking in my experiments as indicative-taking, these verbs may under certain conditions also take subjunctive complements. I discussed the two most prominent of these conditions, which are directive embedding, see (37) above, and negated matrix clauses triggering the 'polarity subjunctive', see (39). As far as my experimental materials are concerned, these two conditions are not relevant. The verbs of this type never embed a directive, and all test sentences are unmodified (= non-negated) declarative clauses. In this environment, the verbs of saying and thinking invariably select for the indicative, even if the speaker rejects the CC in a continuation. Compare (40) with the indicative and matrix negation above, where rejecting the CC was impossible, and non-negated (41) with the standard indicative, where a rejection of the CC is fine in the continuation. I therefore see no reason that the general variability in mood selection affects the interpretation and/or acceptability of the verbs of saying and thinking in my experiments, because their sentence environment invariably calls for the indicative.

(41) María cree que Pedro está enfermo, pero yo no lo creo.
Mary thinks that Peter is-IND ill but I not it think
'Mary believes that Peter is ill but I don't think so.' (Mejías-Bikandi 1994: 899, my gloss and translation)

Let's turn to the two verb types that I pair with the subjunctive in my experiments. These are emotives on the one hand, and the somewhat heterogeneous group comprising desideratives and verbs with implicit negation on the other. Recall from Section 3.1.1 that I assume that the former type are factive and the latter type are non-factive. In addition to these assumptions about factivity, I also made the assumption that both of these types share the property of being non-assertive (as their default interpretation). This means that their usual usage in context is such that their CC is not at issue and does not contain new information. The question for these verbs is now (once again) whether these discourse properties are tied to the mood they select. This question becomes more pressing when we consider that both these types of verbs can, in certain circumstances, also occur naturally with the indicative (see e.g. Lope Blanch 1958, Borrego

Nieto et al. 1986 among many others). Thus, like the verbs of saying and thinking, there is variability in mood selection for these two verb types. Unlike for the verbs of saying and thinking, though, syntactic features of the clause (like negation or question formation) do not seem to trigger the change from one mood to another. Instead, the interpretation of the CC changes with the different mood. For the emotives specifically, Quer (2011) argues that the variability in mood is directly connected to presupposition and assertion. He writes:

> "[W]ith embedded subjunctive the embedded proposition is presupposed, i.e. it is presented as old information present in the context of conversation or information that has to be accommodated as background in that context; with the indicative, the embedded proposition is asserted." (Quer 2011: 167)

While this characterisation is partially true, quantitative evidence from Faulkner (2021a,b) paints a more nuanced picture. Faulkner (2021a,b) reports the results of an acceptability study that tests the interdependence of mood selection and discourse newness of the CC. The study contains the emotives that Quer (2011) is concerned with (for this verb type especially see Faulkner 2021b), but also other predicates that most naturally come with the subjunctive, including those of my subjunctive-taking non-factives (Faulkner 2021a). Her results indicate that the use of the indicative with such verbs indeed receives higher ratings in context where the CC is new in discourse. However, the subjunctive receives equally high ratings, regardless of whether the CC is old or new. This latter finding casts doubt on Quer's (2011) characterization of the subjunctive in emotives, since it does not seem to be exclusively associated with discourse old information (= presupposed content), but seems also to be possible with discourse new information (= asserted content).

Faulkner (2021a,b) concludes that the subjunctive thus is be the default mood for emotives, desideratives, and verbs with implicit negation. For my experimental materials, this means that choosing the subjunctive to go with these verbs was the appropriate thing to do, since it likely caused no undesired meaning shifts for these verb types, and is compatible with any interpretation. Consequently I do not expect that the subjunctive leads to low ratings for these verbs.

The last verb type in my experiments are the knowledge and perception verbs, i.e. semi-factives. These commonly select for the indicative (Terrell & Hooper 1974, Meinunger 2017, among many others). For this verb type there seems to be less concern about mood variability in the literature. I report Lleó's (1979) judgments on the matter, since she directly addresses the interface of mood and factivity for this verb type. She provides the three sentences in (42) with the semi-factive verb *saber* 'know'. In a non-negated sentence, this verb naturally takes the indicative, see (42a). Under negation, the most natural choice is the indicative as well, see (42b), while for the subjunctive it is unclear whether it can occur in this environment, see (42c).

(42) a. Sabía que estabas aquí.
knew-1SG that were-2SG here
'I knew that you were here.'

b. No sabía que estabas aquí.
not knew-1SG that were-IND-2SG here
'I didn't know that you were here.'

c. ? No sabía que estuvieras aquí.
not knew-1SG that were-SUBJ-2SG here
'I didn't know that you were here.'

(Lleó 1979: 170, my glosses and translations)

I am not concerned with the overall acceptability of the subjunctive with *saber* in (42b) – what is important to me are the intuitions Lleó (1979) has as on the presuppositions in (42). She writes the following about these sentences:

> "Even supposing that some [semi-factives] would allow the Subjunctive, this does not involve a concomitant suspension of the presupposition. […] These three sentences presuppose their complements: i.e., the speaker commits himself, at the time of the utterance, to the truth of the complement clause." (Lleó 1979: 170)

For my experiments, this means on the one hand that I chose the appropriate (and perhaps only possible) mood for the semi-factives, and on the other hand that even under the assumption that variability exists for these verbs, this variability is unlikely to affect the interpretation of factivity.

All in all, then, the choices I made on the mood selection of the four verb types seem to have been appropriate. The variability that (some of) these types display are no cause for concern for the experiments. For the verbs of saying (and thinking), I showed that some variability in mood selection is triggered by properties of interpretation (i.e. directive readings) or syntax (the 'polarity subjunctive'). These properties are irrelevant for the experimental materials because they contain unmodified declarative clauses only. In such clauses verbs of saying and thinking always take the indicative, regardless of whether the utterer is certain that the CC is true. The two subjunctive-taking types seem to vary more freely in mood. This variation might have to do with factive and assertive (i.e. non-factive) interpretation to some degree. However, the subjunctive seems to be the default mood for these types, which is compatible with all relevant interpretations. The indicative is restricted to certain discourse environments with these verbs. As a result, the subjunctive is the safe choice here.

Therefore, for the four verb types on the whole, the mood with which they appear in my experimental materials should cause no problems. This is not to say that individual verbs within these types might not have been miscategorized. It is not possible for me to provide an in-depth scrutinizing of every single embedding verb in this chapter, but I would like to touch upon some individual verbs to illustrate this potential problem. Specifically, this issue concerns verbs that display variability in mood selection outside of the systematic variability patterns of their respective types. Rivero (1971: 307 f.) lists a number of verbs that fall into this category. The ones from her list that I include in the experiments are *admitir* 'admit', *esperar* 'hope/expect', *negar* 'deny', *sospechar* 'suspect' and *suponer* 'suppose'. These verbs represent three of my four verb types: the indicative-taking non-factives (*sospechar, suponer*), the subjunctive-taking non-factives (*esperar, negar*) and the semi-factives (*admitir*). Rivero (1971) claims that these verbs may select either the indicative or the subjunctive, independently of other features of the clause like negation. She argues that the choice of mood directly relates to presupposition: with the indicative, the speaker signals

commitment to the truth of the CC, while with the subjunctive, there is no such commitment. Rivero (1971) shows that this commitment inference projects out of polar questions, which strongly suggests that this inference is indeed a factive presupposition.

For my experimental materials, this means that *esperar, negar* and *admitir* are in the correct categories: the first two take the subjunctive and are in the non-factive group. *Admitir* comes with the indicative and is classified as factive. The problematic verbs are *sospechar* and *suponer*, which I classed as non-factive, even though they come with the indicative. If Rivero's (1971) description of these verbs is true,[20] they are in the wrong category in my experiments. While I cannot adduce any evidence to the contrary at this point, I will point out the same thing that I pointed out in the discussion on possible inaccuracies in the categorization of the verbs as factive or non-factive (Section 3.1.1): I am at this point not concerned with this problem, since Chapter 4 features and in-depth quantitative examination of all verbs regarding factivity. The results of this examination will show whether *sospechar* and *suponer* indeed should be classed as factive when they take the indicative. The results of this examination will of course not only illuminate these two verbs, but on all verbs in my experiments as far as their factivity.

3.2 Experimental materials and design

In this section, I describe the design, materials, and participant selection of the experiments. As a reminder on the experiments, I reproduce the overview from the introduction to this chapter (see 6 above) in (43). The design, materials, and participant selection were largely identical across the experiments; the only notable exceptions are Experiment 5, which tested less complex materials than those of the other experiments, and Experiment 7, which contained more verb types than the others and therefore more materials and experimental factors. I concentrate on the common features of the experiments in this section, any differences between experiments are discussed in the relevant sub-sections.

(43) **Overview over the Experiments**

1. Gapping vs. stripping under indicative-taking non-factives
2. Gapping vs. stripping under subjunctive-taking true factives
3. Indicative-taking non-factives vs. subjunctive-taking true factives in gapping
4. Full clause equivalents of Experiment 3
5. Indicative-taking semi-factives vs. subjunctive-taking non-factives in gapping I
6. Indicative-taking semi-factives vs. subjunctive-taking non-factives in gapping II
7. All four verb types in gapping

All experiments except for Experiment 7 had a 2×2 factorial design, with one fixed factor being the presence/absence of the complementizer *que* in the second conjunct. The other factor was either the type of elliptical structure (gapping vs. stripping, Experiments 1 and 2) or factivity (non-factive vs. factive, Experiments 3 to 6). Thus, target sentences in these experiments came in four experimental conditions. Example (44) shows one such target sentence from Experiment 1 in all four conditions. The polarity particle serving as a remnant in the stripping conditions was invariably *también* 'too'. The sentences in (45) and **(46)**

[20] Rivero's (1971) judgments do not seem to be overly robust in general: Klein (1977) and Lleó (1979) disagree with some aspects of Rivero's (1971) basic introspective data. These disagreements do not affect the list of verbs in question though.

show a sample item from Experiment 3, which only tested gapping, but manipulated the embedding verbs. Sentence (45) shows the non-factive conditions with the indicative-taking verb of thinking *cree* 'thinks'. Sentence (46) shows the factive conditions with the true factive, subjunctive-taking verb *desaprueba* 'disapproves'. Experiments 5 and 6 had the same design, but tested different verb types (semi-factives and subjunctive-taking non-factives). See their respective sections for sample items.

(44) **Target sentence Experiment 1, all four conditions**
Santiago cree que el rey fuma cada día y {que / Ø} la reina, {a veces / también}.
Santiago thinks that the king smokes-IND every day and that the queen at times too
'Santiago thinks that the king smokes every day and that the queen smokes sometimes / that the queen smokes every day, too.'

(45) **Target sentence Experiment 3, non-factive conditions**
Santiago cree que el rey fuma cada día y {que / Ø} la reina, a veces.
Santiago thinks that the king smokes-IND every day and that the queen at times
'Santiago thinks that the king smokes every day and that the queen smokes sometimes.'

(46) **Target sentence Experiment 3, factive conditions**
Santiago desaprueba que el rey fume cada día y {que / Ø} la reina, a veces.
Santiago disapproves that the king smokes-SUBJ every day and that the queen at times
'Santiago disapproves of the fact that the king smokes every day and that the queen smokes sometimes.'

The structure of the target sentences was as follows. The matrix subjects were personal names (like *Santiago* in 44 to 46), definite DPs like *el médico* 'the doctor' (see Section 3.3.6 below for an example), or a definite DP containing a personal name like *el diputado Hernández* 'Congressman Hernández'. All matrix subjects were in the third person singular. The embedded clauses consisted of intransitive verbs like *fuma* 'smokes' in (44) to (46), or verbs that take the reflexive pronoun *se* like in *se acuestan* '(they) go to bed'. The embedded subjects in both conjunct were definite DPs in the third person singular or plural. Additionally to the subjects, the embedded clauses featured adverbials (e.g. *cada día* 'every day' in 44 to 46). In sentences with stripping, this concerned the first conjunct only; in the elliptical conjunct the subject was the sole remnant constituent. In all other sentences, both embedded conjuncts contained an adverbial. In sentences with gapping, the subject and the adverbial were thus the two remnant constituents.

The complete experimental materials consisted of 24 embedded coordinate clauses like just described, and the 12 embedding verbs of each of the four types. This meant that each verb occurred twice, but with two different embedded clauses, which resulted in a total of 96 verb-clause combinations. Each of these combination came in one version with *que* and another version without *que*, which made a total of 192 items. In all experiments (except 5 and 7), a subset of these items was distributed across four lists, containing each embedded clause once in one of the experimental conditions (= Latin square). There were 36 additional filler items on each list, which meant that each list contained a total of 60 sentences to be judged by the participants. (Each participants only judged one list each.) The target sentences (both experimental and filler items) were introduced by a context consisting of one or two sentences, which described a situation in which the CC was made plausible. Further, the context prepared the desired (non-)factive interpretation of the sentence so that the CC could plausibly be interpreted as a matter of discussion for non-factive verbs, and as a fact for factive verbs. In the majority of the cases, an embedding

clause received the same context with all its different embedding verbs, since the contexts were generic enough to allow for non-factive and factive interpretations simultaneously.

In stripping without *que*, the position of the remnant subject was potentially ambiguous between the main clause and the embedded clause. For example, in (47) the elliptical conjunct may either receive the desired embedded interpretation *The queen smokes every day, too*, or the non-desired matrix interpretation *The queen also thinks that the king smokes every day*. To ensure that participants got the intended embedded interpretation, I constructed the items such that the necessary semantic parallelism with the remnant subject more plausibly obtained for the embedded subject than for the matrix subject. In (47), the remnant subject *queen* is more likely to be interpreted as correlating with the embedded subject *king* than with the matrix subject *Santiago*. Furthermore, the context made this interpretation more prominent. Finally, each target sentence was followed by a verification question on the following page of the questionnaire, so as to guarantee that participants had understood the target sentence as intended.[21]

(47) **Target sentence Experiment 1, stripping without *que***
Santiago cree que el rey fuma cada día y la reina, también.
Santiago thinks that the king smokes every day and the queen too
'Santiago thinks that the king smokes every day and the queen, too.'

The 36 filler items contained a mix of sentences with clausal embedding, non-elliptical coordinate main clauses, coordinate main clauses with gapping or stripping, and "regular" main clauses with none of these features. As a result, the set of filler items contained all the crucial features that each experimental item had (= clausal co- and subordination plus ellipsis), but none of the filler items had all of these features simultaneously. The design of the filler items was such that the expected acceptability ratings of one half would be high, and the other half low. This provided a baseline of acceptability and unacceptability.[22] The unacceptable controls were structurally unacceptable or semantically incongruent sentences. The general set of control items remained constant across the experiments, with minor adjustments of individual control items between experiments. I made sure that none of the controls that featured clausal embedding contained any of the embedding verbs of the test sentences in their respective experiment.

The experiments were web-based questionnaires hosted on SoSciSurvey.de (Leiner 2024). Experimental items and fillers were presented one-by-one on separate pages. These pages contained the test sentence, preceded by its context, and followed by a continuous horizontal scale on which participants gave their judgments. The right extreme of this scale read *totalmente natural* 'totally natural', and the left extreme read *totalmente forzado* 'totally unnatural' (literally 'totally forced'). Participants dragged a slider between the extremes, corresponding to their judgment of naturalness. These judgments were then coded with integers from 1 (totally unnatural) to 100 (totally natural). Each page containing context, sentence, and scale was followed by a page containing the verification question on that sentence.

[21] Only a low percentage of questions received incorrect answers: for stripping, this was the case for 3% and 6% in Experiment 1 and 2, respectively; for gapping, this was the case for 7% and 4% in Experiment 1 and 2, respectively (all numbers after participant exclusion, see below). The other experiments had similar percentages of incorrect answers. I take these low numbers to indicate that structural ambiguity was no major issue.

[22] As we will see in the results, the mean ratings for the unacceptable controls in all experiments seem to be too high to indicate outright unacceptability. (This assessment is based entirely on my introspective idea of how low ratings for outright unacceptability should have been.) I do not consider this to be a problem, since these ratings still provide the desired baseline. If an experimental condition is rated higher than the unacceptable controls, I still may conclude that this condition is not unacceptable.

The variety of Spanish investigated in my study was European Spanish.[23] I used the online platforms Prolific.com and Amazon Mechanical Turk (mturk.com) for recruiting participants. A pre-survey questionnaire identified the variety of Spanish that participants spoke as well as their use of other languages. Data from speakers of Latin American Spanish (up to 33% per experiment) and from (active, habitual) bilingual speakers were excluded. The bilingual exclusion criterion applied to up to 37% of participants per experiment, and largely concerned speakers of other Iberian languages such as Catalan, Galician, and Basque. Despite my best efforts to only include native speakers of Spanish via the pre-screening options on Prolific and MTurk, some participants indicated that their native language was not Spanish. Data from these participants did not enter the analysis.

Further, participants who did not differentiate between acceptable and unacceptable fillers (as determined by the non-significant result of a one-sided t-test for each participant), or who answered more than 20% of the verification questions for the filler items incorrectly were also excluded from data analysis (on average 8% and 6% of participants per experiment, respectively). Table 3.2 summarizes these numbers for each experiment. It shows the total number of participants who finished their questionnaire, the number of participants affected by each exclusion criterion, and the number of remaining participants whose data entered the analysis. In all the experiments, some criteria excluded the same participants simultaneously. For example, non-native speakers of Spanish likely also gave undifferentiated ratings for the high and low acceptability controls, and likely answered more than 20% of verification questions incorrectly, which resulted in their being excluded for each of these three different criteria. For this reason, the total number of participants in each column is lower than the sum of the other lines.

Table 3.2: Total number of participants, total numbers and percentages of participants affected by each exclusion criterion, and total number and percentage of participants remaining in the data per experiment. Several participants were excluded by multiple criteria in each experiment. Percentages are from the Total (first row) of the respective column.

	Experiments						
	1	2	3	4	5	6	7
Total	58	43	65	71	61	70	108
Non-Spanish	0 (0%)	1 (2%)	1 (2%)	2 (3%)	4 (7%)	6 (9%)	7 (6%)
Latin American	19 (33%)	2 (5%)	8 (12%)	13 (18%)	9 (15%)	11 (16%)	15 (14%)
Multilingual	18 (31%)	12 (28%)	25 (38%)	25 (35%)	18 (30%)	26 (37%)	30 (28%)
Control ratings	5 (9%)	4 (9%)	1 (2%)	6 (8%)	7 (11%)	6 (9%)	9 (8%)
Control questions	3 (5%)	2 (5%)	2 (3%)	3 (4%)	3 (5%)	8 (11%)	7 (6%)
Remaining	27 (47%)	27 (63%)	32 (49%)	35 (49%)	30 (49%)	30 (43%)	60 (56%)

For the statistical analysis I used `R` (R Core Team 2023) with the packages `afex` (Singmann et al. 2022) and `lmer` (Kuznetsova et al. 2017) to fit linear mixed models for the data of each experiment as well as for cross-experiment comparisons. Since Experiments 1 to 6 had similar 2×2 factorial designs, their models had similar fixed effects structures: the first fixed factor was the presence/absence of the complementizer,

[23] Although I have no immediate reason to believe that varieties differ w.r.t. the acceptability of embedded gapping with *que*, related structures do behave differently across varieties. For instance, Martínez Vera (2020) discusses differences between European and Latin American Spanish w.r.t. recomplementation, a structure which I touch upon in Section 3.5. I decided not to explore potential differences between different varieties in my study to keep the experimental design simple. (Potential) variation is an issue that needs to be explored in future research. Note, however, that the corpus tokens presented in Chapter 2.5 stem from Latin American sources, which suggests that embedded gapping with *que* is not unacceptable in this variety.

and the second factor differed between the experiments (ellipsis type for Experiments 1 and 2, factivity for Experiments 3 to 6). All models included the interaction of both fixed factors. Factors were sum-to-zero contrast coded to facilitate the interpretation of interaction effects. The models had a random structure that included participants and the contents of the embedded clauses (≈ items). In line with Barr et al. (2013), the model with the maximal random effects structure that converged (and was not a singular fit) formed the basis for statistical evaluation. I used the R-package `ggplot2` (Wickham 2016) to generate the figures. The numbers in the results tables are based on the `summarySEwithin()`-function of the R-package `Rmisc` (Hope 2022). This function removes inter-subject variability.

3.3 Results

This section contains the results of the seven experiments, with one subsection per experiment. Additionally, there are two further subsections, one after Experiment 2 and one after Experiment 6, which contain short interim discussions.

3.3.1 Experiment 1

Experiment 1 investigated gapping and stripping under indicative-taking non-factive verbs, i.e. verbs of saying and thinking. For a sample item in all conditions, see (48). Each participant judged each embedded clause once and each embedding verb twice (in different experimental conditions). In total, 58 participants completed the survey. The data of 27 of them (16 female, 11 male, ages 20-60, median age 33) entered the analysis. Table 3.3 and Figure 3.1 show the results for the four conditions.

(48) **Target sentence Experiment 1** (same as example 44 above)
Santiago cree que el rey fuma cada día y {que / Ø} la reina, {a veces / también}.
Santiago thinks that the king smokes-IND every day and that the queen at times too
'Santiago thinks that the king smokes every day and that the queen smokes sometimes / that the queen smokes every day, too.'

Table 3.3: Experiment 1 (indicative non-factives): Mean ratings per experimental condition and control items (*N* = number of observations, *SD* = standard deviation, *SE* = standard error, *CI* = 95% confidence interval)

Ellipsis Type	Complementizer	N	Mean Rating	SD	SE	CI
Gapping	Ø	150	80.06	22.88	1.87	3.69
Gapping	*que*	148	75.00	25.38	2.09	4.12
Stripping	Ø	157	73.25	21.32	1.70	3.36
Stripping	*que*	159	73.56	22.04	1.75	3.45
High acceptability controls		468	89.83	24.39	1.13	2.22
Low acceptability controls		431	29.03	39.27	1.89	3.72

Gapping sentences without *que* received slightly higher ratings than gapping sentences without *que*. Within the stripping sentences, there is hardly a difference between mean ratings. In the statistical model, neither the effect of type of ellipsis ($b = 2.29, SE = 1.57, t = 1.46, p = 0.16$), nor the presence of *que*

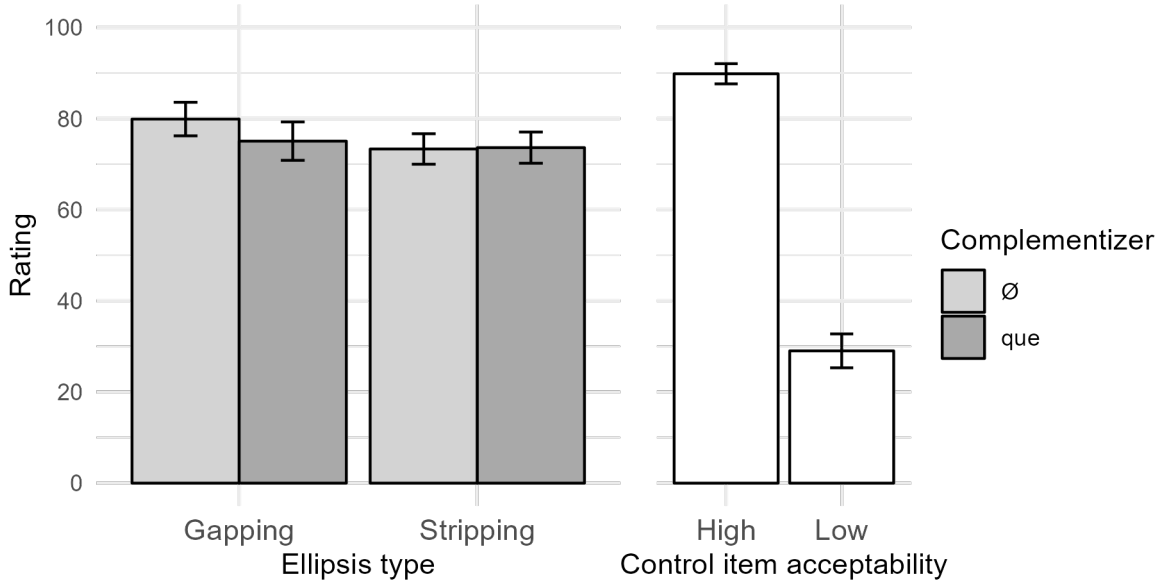


Figure 3.1: Results for Experiment 1 (indicative non-factives): Mean ratings with 95% confidence interval.

($b = 1.22, SE = 1.03, t = 1.19, p = 0.25$), nor the interaction between these two fixed factors ($b = 1.23, SE = 0.73, t = 1.68, p = 0.09$) was significant. Since the interaction effect approached significance at $p = 0.09$, I split the experimental data by ellipsis type into a gapping set and a stripping set, and fit a linear model each. For the gapping sentences, the effect of *que* was not significant ($b = 2.43, SE = 1.31, t = 1.85, p = 0.08$). The effect of *que* in the stripping data set was clearly not significant either ($b = -0.01, SE = 0.96, t = -0.01, p = 0.99$).

These results provide evidence that *que* is acceptable in Spanish globally embedded gapping and stripping under verbs of saying and thinking. In fact, this structure seems to be just as acceptable overall as its counterpart without *que*. Further, there does not seem to be a meaningful difference between the ellipsis types gapping and stripping.

3.3.2 Experiment 2

Experiment 2 tested gapping and stripping with true factive embedding verbs, i.e. emotive verbs, which select for the subjunctive. Otherwise, the design was equivalent to that of Experiment 1. A sample of a target sentence is in (49). It is identical to the sample item for Experiment 1 in (48), except for the embedding verb (*desaprueba* 'disapproves' instead of *cree* 'thinks') and the mood of the embedded verb (subjunctive *fume* 'smokes' instead of indicative *fuma*).

(49) **Target sentence Experiment 2**
Santiago desaprueba que el rey fume cada día y {que / Ø} la reina, {a veces / también}.
Santiago disapproves that the king smokes-SUBJ every day and that the queen at times too
'Santiago disapproves of the fact that the king smokes every day and that the queen smokes sometimes / that the queen smokes every day, too.'

43 participants completed the survey. The data of 27 of them (6 female, 21 male, ages 19-51, median age 33) entered the analysis. Table 3.4 and Figure 3.2 show the results. In the results, there is a notable difference between clauses with and without *que* within the gapping ratings. Clauses without *que* are rated as more acceptable than those with *que*. Within the stripping ratings, there is virtually no difference.

Table 3.4: Experiment 2 (subjunctive true factives): Mean ratings per experimental condition and control items (*N* = number of observations, *SD* = standard deviation, *SE* = standard error, *CI* = 95% confidence interval)

Ellipsis Type	Complementizer	N	Mean Rating	SD	SE	CI
Gapping	Ø	150	71.09	21.42	1.75	3.46
Gapping	*que*	148	58.49	25.81	2.12	4.19
Stripping	Ø	142	65.75	23.38	1.96	3.88
Stripping	*que*	153	65.48	21.09	1.71	3.37
High acceptability controls		455	89.23	26.16	1.23	2.41
Low acceptability controls		448	31.84	46.42	2.19	4.31

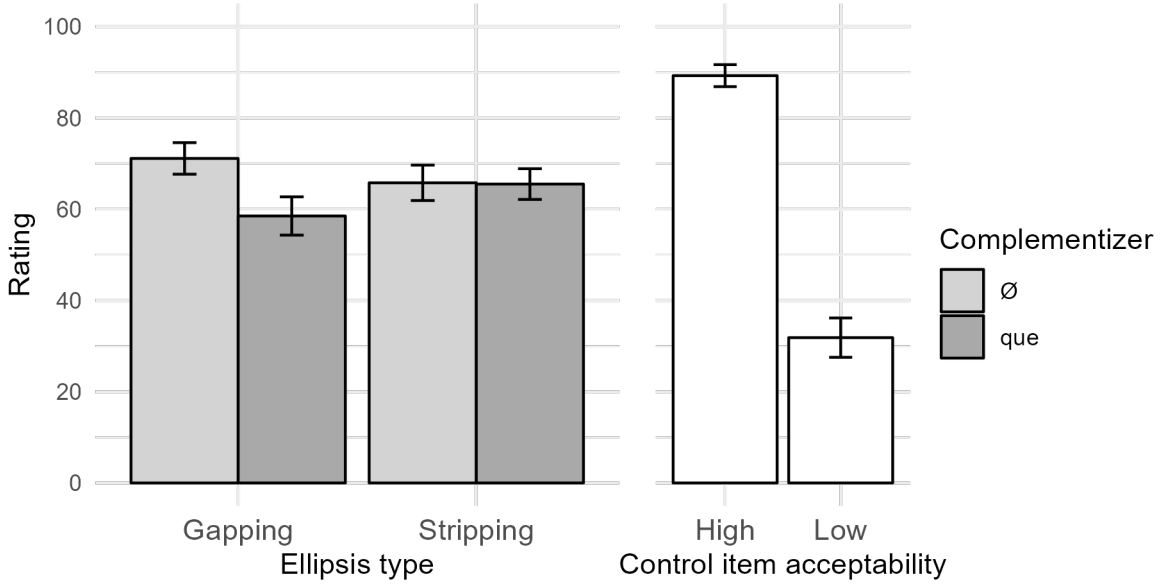


Figure 3.2: Results for Experiment 2 (subjunctive true factives): Mean ratings with 95% confidence interval.

The statistical model confirmed these observations. It indicated no significant effect of the type of ellipsis ($b = -0.41, SE = 0.90, t = -0.45, p = 0.66$). The effect of *que* was significant ($b = 3.28, SE = 1.47, t =$

2.24, $p = 0.03$) and there was a significant interaction of the two fixed factors ($b = 3.14, SE = 3.18, t = 2.66, p = 0.01$). To resolve the interaction, I split the data into a gapping set and a stripping set and conducted single comparisons. The model for the gapping data indicated a significant effect of *que* ($b = 6.28, SE = 1.21, t = 5.17, p < 0.001$), but the model for the stripping data indicated no significant effect ($b = 0.30, SE = 1.42, t = 0.21, p = 0.84$).

Despite the significantly worse ratings for *que* in gapping, this condition does not appear to be outright unacceptable. The mean rating for these clauses is still well over 50. Furthermore, ratings for this conditions are clearly higher than those for the unacceptable control items, see Figure 3.2. This indicates that participants still made a categorical difference between gapping clauses with *que* under factives and low acceptability controls. Figure 3.2 also indicates that none of the experimental conditions patterned with the acceptable controls either. I tentatively take this finding to be an effect of the complexity of the test sentences, which all included clausal embedding, clausal coordination and ellipsis at the same time. The acceptable control items were less complex in structure.

3.3.3 Comparing Experiments 1 and 2

Comparing the two experiments, I conclude that *que* is acceptable in globally embedded gapping and stripping in Spanish. None of the relevant conditions approach ratings that are near the low acceptability control items. Moreover, I tentatively conclude that the acceptability of *que* in gapping is constrained by the factivity of the embedding verb: I found no significant difference between clauses with and without *que* under non-factives (Experiment 1), but did so under factives (Experiment 2). For stripping clauses I found no significant effect of *que* in either experiment, which suggests that *que* in this ellipsis type is not sensitive to the factivity of the embedding verb. While this difference between the ellipsis type is certainly interesting, I will leave the stripping data aside and concentrate on gapping. (Remember that I only included stripping as a baseline for gapping in the first place. Since there seems to be no general difference between the ellipsis types in terms of acceptability, stripping has "done its duty" for the purposes of this study.)

Another observation is that the ratings across all four experimental conditions are notably lower in Experiment 2 than Experiment 1. A linear mixed effects model of the combined experimental items revealed a significant difference between experiments ($b = -5.42, SE = 2.67, t = -2.03, p = 0.048$). In contrast to the experimental items, the ratings for the filler items, which were identical in the two experiments, were relatively constant across experiments, see Table 3.5. I fitted a linear mixed effects model of the combined filler data from Experiments 1 and 2. It contained experiment (1 vs. 2) and predicted acceptability (high vs. low), as well as the interaction term as fixed factors. The model revealed no significant effect for the different experiments ($b = 0.45, SE = 1.17, t = 0.40, p = 0.70$) or the interaction of acceptability and experiment ($b = 1.11, SE = 1.02, t = 1.09, p = 0.28$). Unsurprisingly, low acceptability controls received significantly lower ratings than high acceptability controls ($b = -29.21, SE = 2.53, t = -11.54, p < 0.001$). The fact that I found significantly lower ratings for the experimental items between the experiments, but not for the filler items, leads me to conclude that embedded elliptical clauses are generally less acceptable under factives than under non-factives. To corroborate this conclusion, and in order to see whether the

Table 3.5: Comparison of control item ratings of Experiments 1 and 2

Experiment	Acceptability	Mean Rating	SD
1	High	90.20	17.44
2	High	89.16	19.89
1	Low	29.25	29.67
2	Low	32.28	35.30

experimental results for gapping would replicate, Experiment 3 explored the difference between non-factives and factives again.

3.3.4 Experiment 3

Experiment 3 examined the gapping materials of Experiments 1 and 2. Every participant saw every embedding verb and every embedded clause once, either with a non-factive embedding verb (and indicative mood on the embedded verb) or with a factive verb (and subjunctive mood on the embedded verb). For sample items, see (50) and (51), respectively, which are repeated from Section 3.2 above.

(50) **Target sentence Experiment 3, non-factive conditions** (same as gapping conditions of Experiment 1)
Santiago cree que el rey fuma cada día y {que / Ø} la reina, a veces.
Santiago thinks that the king smokes-IND every day and that the queen at times
'Santiago thinks that the king smokes every day and that the queen smokes sometimes.'

(51) **Target sentence Experiment 3, factive conditions** (same as gapping conditions of Experiment 2)
Santiago desaprueba que el rey fume cada día y {que / Ø} la reina, a veces.
Santiago disapproves that the king smokes-SUBJ every day and that the queen at times
'Santiago disapproves of the fact that the king smokes every day and that the queen smokes sometimes.'

65 participants completed the survey. The data of 32 of them (8 female, 24 male, ages 18-50, median age 32) entered the analysis. The results of Experiment 3 are shown in Table 3.6 and Figure 3.3. The linear model indicated significant effects for both experimental factors. Sentences without *que* were rated significantly better than sentences with *que* ($b = 3.45, SE = 0.92, t = 3.75, p = 0.001$). Factives were rated significantly worse than non-factives ($b = -2.52, SE = 0.99, t = -2.56, p = 0.02$). The interaction of the fixed effects was significant as well ($b = 2.45, SE = 0.77, t = 3.19, p = 0.005$).

To resolve the interaction, and in order to explore the difference I observed between experiments 1 and 2, I partitioned the data twice, once by verb type and once by presence/absence of *que* to fit linear models to these sub-sets. Due to the resulting multiple comparisons within the same data, I report Bonferroni-Holm-corrected p-values for the four tests I conducted. First, I split the data into a non-factive and a factive set and fitted a linear model for each set. For the non-factive data set, the model indicated no significant effect of *que* on the ratings ($b = 1.00, SE = 0.93, t = 1.07$, corrected $p = 0.6$). For the factive data set, the model did find a significant effect of *que* ($b = 5.87, SE = 1.19, t = 4.96$, corrected $p < 0.001$). This finding leads me to conclude that it is only under factives that *que* causes a degradation in acceptability.

The split by presence/absence of *que* yielded the following result: the model for the Ø-set indicated no significant effect of factivity ($b < 0.01, SE = 0.87, t = 0.01$, corrected $p = 0.99$). The model for the *que*-set

Table 3.6: Experiment 3 (gapping with indicative non-factives and subjunctive true factives): Mean ratings per experimental condition and control items (*N* = number of observations, *SD* = standard deviation, *SE* = standard error, *CI* = 95% confidence interval)

Verb type	Complementizer	N	Mean Rating	SD	SE	CI
Non-factive	Ø	174	82.10	18.99	1.44	2.84
Non-factive	*que*	174	80.42	20.88	1.58	3.12
True factive	Ø	182	82.05	18.90	1.40	2.76
True factive	*que*	170	70.14	27.70	2.12	4.19
High acceptability controls		555	91.93	23.87	1.01	1.99
Low acceptability controls		536	34.33	43.68	1.89	3.71

did indicate a significant effect for factivity ($b = -4.86, SE = 1.53, t = -3.17$, corrected $p = 0.015$).

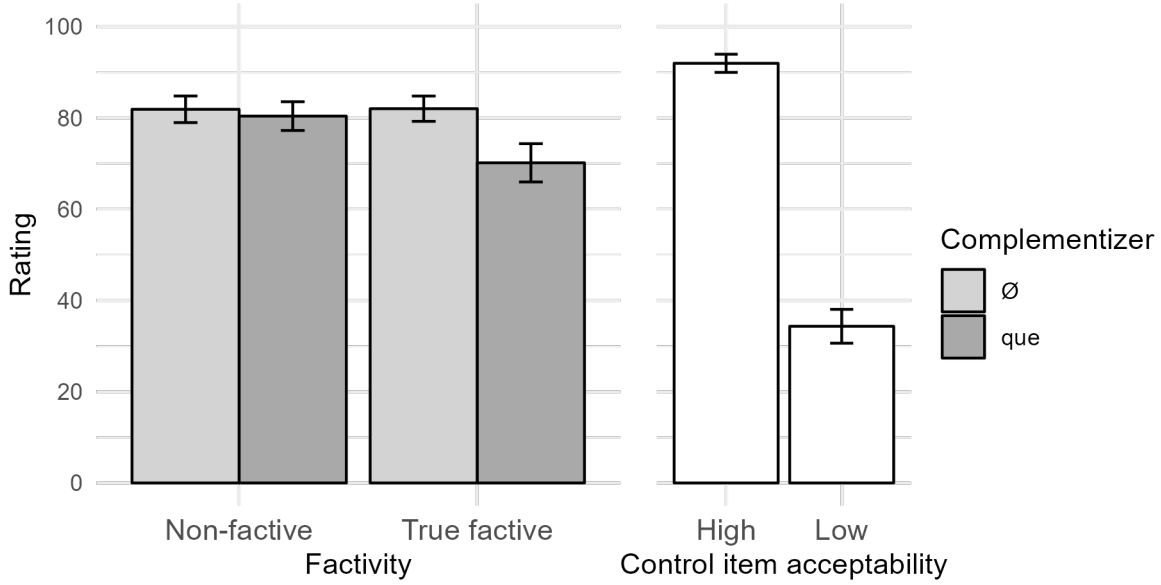


Figure 3.3: Results of Experiment 3 (indicative non-factives vs. subjunctive true factives): Mean ratings with 95% confidence interval.

Thus, the results of Experiments 1 and 2 were replicated: Clauses with *que* are significantly degraded with factive verbs, but not with non-factives. What was not replicated was the overall difference between non-factives and factives from Experiments 1 and 2. Although the statistical model indicates a significant difference between non-factives and factives in Experiment 3 as well, the single comparison shows that this is due to only one degraded condition (factives with *que*), and not a general property of non-factives vs. factives.

3.3.5 Experiment 4

Experiment 4 examined the full-clause equivalents of Experiment 3. The materials were thus identical to those of Experiment 3, with the exception that they included overt verbs in the second conjunct of each target item. This resulted in target sentences like (52) and (53) for the non-factive and factive conditions, respectively. The results are in Table 3.7 and Figure 3.4. All four conditions receive relatively high ratings.

The conditions with *que* are very slightly preferred to the conditions without *que*, for both the non-factives and factives. However, the linear model indicated that this effect of *que* was not significant ($b = -1.08, SE = 0.73, t = -1.48, p = 0.16$). The effect of the factivity ($b = -0.24, SE = 1.13, t = -0.21, p = 0.83$) and the interaction effect ($b = -0.13, SE = 0.65, t = -0.21, p = 0.84$) were not significant either. The *que*-effect within the factives that I observed for gapping in Experiment 3 thus did not replicate in full clauses.

(52) **Target sentence Experiment 4, non-factive conditions**
Santiago cree que el rey fuma cada día y {que / Ø} la reina fuma a
Santiago thinks that the king smokes-IND every day and that the queen smokes-IND at
veces.
times
'Santiago thinks that the king smokes every day and that the queen smokes sometimes.'

(53) **Target sentence Experiment 4, factive conditions**
Santiago desaprueba que el rey fume cada día y {que / Ø} la reina fume
Santiago disapproves that the king smokes-SUBJ every day and that the queen smokes-SUBJ
a veces.
at times
'Santiago disapproves of the fact that the king smokes every day and that the queen smokes sometimes.'

Table 3.7: Experiment 4 (full clause equivalents of Experiment 3): Mean ratings per experimental condition and control items (*N* = number of observations, *SD* = standard deviation, *SE* = standard error, *CI* = 95% confidence interval)

Verb type	Complementizer	N	Mean Rating	SD	SE	CI
Non-factive	Ø	198	80.55	21.31	1.51	2.99
Non-factive	*que*	199	82.67	20.49	1.45	2.86
True factive	Ø	192	79.79	22.83	1.65	3.25
True factive	*que*	192	82.37	21.09	1.52	3.00
High acceptability controls		598	89.05	27.56	1.13	2.21
Low acceptability controls		581	30.21	40.69	1.69	3.32

3.3.6 Experiment 5

Experiment 5 investigated two new sets of embedding verbs: indicative-taking semi-factives, and another set of non-factives, which select for the subjunctive. The new non-factives were desideratives and verbs with implicit negation. See (54) and (55) for a sample with the non-factive negation verb *niega* 'denies' and the semi-factive knowledge verb *ignora* 'ignores', respectively.

(54) **Target sentence Experiment 5, non-factive conditions**
Santiago niega que el rey fume cada día y {que / Ø} la reina, a veces.
Santiago denies that the king smokes-SUBJ every day and that the queen at times
'Santiago denies that the king smokes every day and that the queen smokes sometimes.'

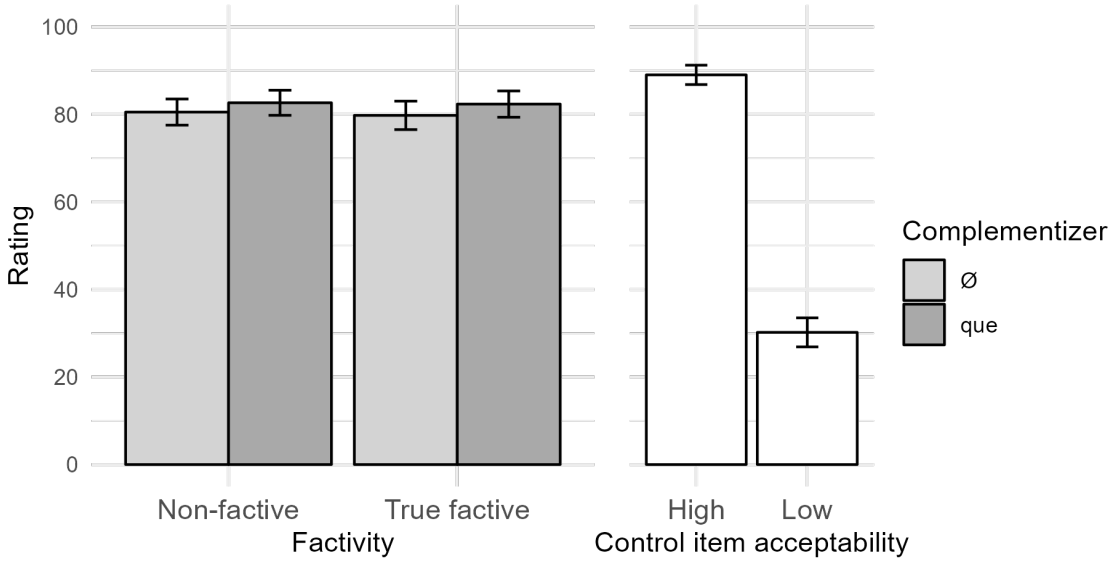


Figure 3.4: Results of Experiment 4 (full clause equivalents of Experiment 3): Mean ratings with 95% confidence interval

(55) **Target sentence Experiment 5, factive conditions**
El médico ignora que el rey fuma cada día y {que / Ø} la reina, a veces.
the doctor ignores that the king smokes-IND every day and that the queen at times
'The doctor ignores that the king smokes every day and that the queen smokes sometimes.'

I changed the 'old' embedding verbs for the 'new' ones such that each of the 24 new verbs came with one embedding clause. This experiment thus only contained half the amount of the materials as each of the preceding experiments, which all paired each embedding verb with two embedding clauses. Consequently, the items in Experiment 5 were divided onto two lists instead of four. Some of the lexical material of some of the embedded clauses was slightly adjusted so as to fit the lexical meaning of the embedding verb. Likewise, contexts were adjusted so as to facilitate participants' interpretation of the lexical material of the embedded clauses, of the lexical meaning of the new embedding verb, and the intended non-factive or factive interpretation. Some of the matrix subjects were adjusted as well, for instance the definite DP *el médico* 'the doctor' replaced the proper name *Santiago* in (55), because the context had been changed from a family discussion setting to a medical situation.

61 participants completed the questionnaire. The data of 30 of them (12 female, 18 male, ages 22-50, median age 34) entered the analysis. The results are in Table 3.8 and Figure 3.5. All conditions received relatively high ratings. For both verb types, the version with *que* was rated worse than the version without *que*. For the factives, this difference was larger than for the non-factives. The statistical model indicated a significant effect of the complementizer ($b = 6.23, SE = 2.01, t = 3.10, p < 0.01$), and a significant interaction effect ($b = 1.76, SE = 0.72, t = 2.45, p = 0.02$). The factivity effect was not significant ($b = 1.16, SE = 1.51, t = 0.77, p = 0.45$). To resolve the interaction, I split the data into a non-factive and a factive sub-set, and fit a linear model each. The model for the non-factives indicated that the effect of *que* was not significant, but albeit at the alpha-level of .05 ($b = 4.50, SE = 2.19, t = 2.05, p = 0.05$). The model for the factives on the other hand indicated that the effect of *que* was significant ($b = 7.84, SE = 2.42, t =$

3.24, $p < 0.01$). Since the *que*-effect for the non-factives approached significance at $p = 0.05$, I decided to test these verbs again in another experiment (Experiment 6)

Table 3.8: Experiment 5 (indicative semi-factives and subjunctive non-factives): Mean ratings per experimental condition and control items (*N* = number of observations, *SD* = standard deviation, *SE* = standard error, *CI* = 95% confidence interval)

Verb type	Complementizer	N	Mean Rating	SD	SE	CI
Non-factive	Ø	166	82.78	24.55	1.91	3.76
Non-factive	*que*	170	73.77	26.10	2.00	3.95
Semi-factive	Ø	168	88.42	19.73	1.52	3.00
Semi-factive	*que*	158	72.77	27.20	2.16	4.27
High acceptability controls		522	94.17	20.80	0.91	1.79
Low acceptability controls		513	37.86	47.92	2.12	4.16

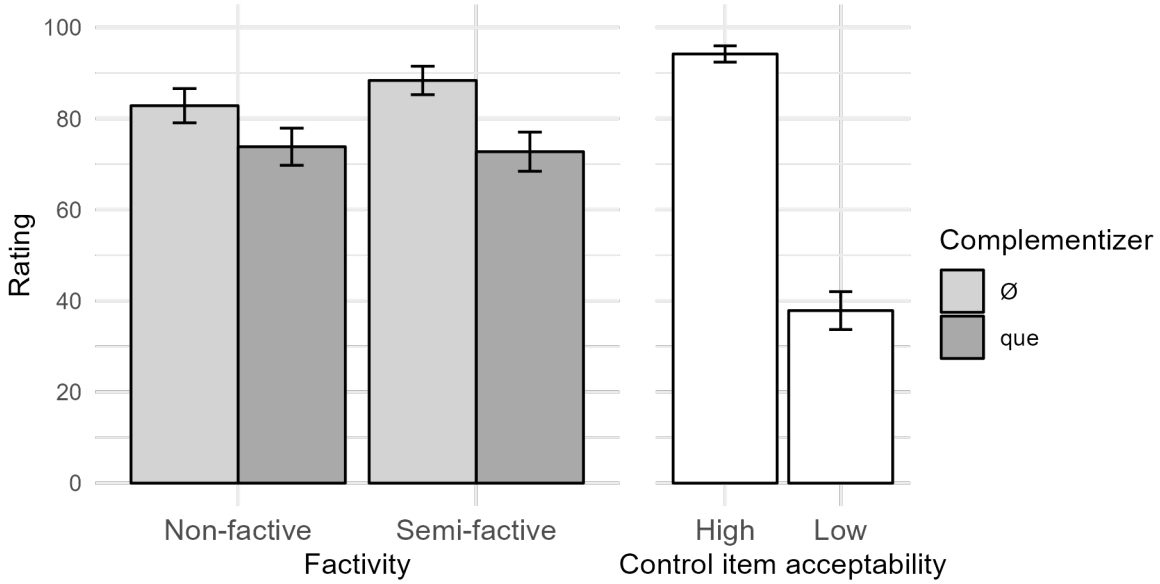


Figure 3.5: Results of Experiment 5 (subjunctive non-factives vs. indicative semi-factives): Mean ratings with 95% confidence interval

3.3.7 Experiment 6

Experiment 6 tested the verb types of Experiment 5 again (i.e. indicative-taking semi-factives and subjunctive-taking non-factives). The difference between these two experiments was that Experiment 6 doubled the number of experimental items by pairing each embedding verb with another embedded clause, thus having each verb with two embedded clauses. This way, the design of Experiment 6 copied that of Experiment 3, the only difference being the choice of embedding verbs and some minor adjustments of the items. For details on these adjustments and sample items, see Section 3.3.6 on Experiment 5.

70 participants completed the survey. The data of 30 of them (10 female, 19 male, 1 n.a., ages 18-45, median age 29) entered the analysis. Table 3.9 and Figure 3.6 show the results. Like in Experiment 5,

ratings for *que* are lower than ratings with *que* for both verb types. However, this time the difference is larger for the non-factives than for the factives, which is the opposite of what we saw in Experiment 5.

The statistical model yielded the following results. The difference between semi-factives and non-factives is not significant ($b = 1.10, SE = 1.03, t = 1.07, p = 0.30$). Clauses with *que* are rated significantly worse than clauses without *que* ($b = 4.12, SE = 1.72, t = 2.93, p < 0.01$). Despite appearances, the interaction of the two fixed effects is not significant ($b = -1.05, SE = 0.66, t = -1.59, p = 0.11$). I will discuss these results and compare them to those of Experiment 5 in the following section.

Table 3.9: Experiment 6 (indicative semi-factives and subjunctive non-factives): Mean ratings per experimental condition and control items (*N* = number of observations, *SD* = standard deviation, *SE* = standard error, *CI* = 95% confidence interval)

Verb type	Complementizer	N	Mean Rating	SD	SE	CI
Non-factive	Ø	162	78.04	20.79	1.63	3.23
Non-factive	*que*	164	67.16	25.12	1.96	3.87
Semi-factive	Ø	167	78.13	20.58	1.59	3.14
Semi-factive	*que*	163	71.59	22.76	1.78	3.52
High acceptability controls		505	86.13	27.45	1.22	2.40
Low acceptability controls		499	32.10	42.85	1.92	3.77

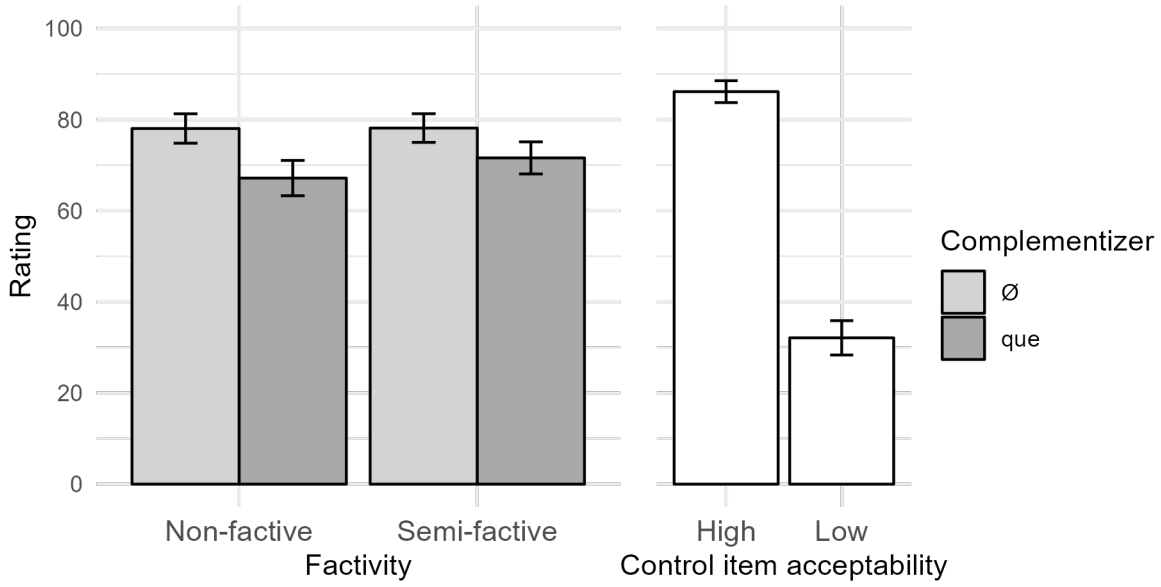


Figure 3.6: Results of Experiment 6 (subjunctive non-factives vs. indicative semi-factives): Mean ratings with 95% confidence interval

3.3.8 Discussion of Experiments 1 to 6

Experiments 1 to 6 provide relevant data for the three hypotheses that I formulate at the beginning of this chapter. I repeat these hypotheses here in (56). I have already discussed Hypothesis 2 on the impact of the type of ellipsis in Section 3.3.3 above. For this reason I will focus on the other two hypotheses.

(56) **Hypotheses**: The acceptability of *que* in global embedding is constrained by
1. The type of embedding verb (non-factive vs. semi-factive vs. true factive)
2. The ellipsis type (gapping vs. stripping)
3. Ellipsis as such (gapping vs. full clauses)

Hypothesis 3 was that the sensitivity to *que* is different in gapping than in full clauses. A comparison of the results of Experiments 3 and 4 suggests that this is indeed the case: in gapping, there is a degradation of *que* under true factives (Experiments 2 and 3). In the corresponding full clauses, there is no such degradation (Experiment 4). This observation strongly suggests that the degrading effect of *que* under true factives is indeed particular to gapping, and not a general property of embedded coordinate clauses with these verbs. Of course, the lack of a statistically significant *que*-effect for true factives in full clauses might be coincidental. However, this verb type consistently shows a significant effect in the two experiments that tested them in gapping (Experiments 2 and 3, see also Experiment 7 below for further confirmation). Since the overall design and materials of these experiments was equivalent, I am confident that the measurement in Experiment 4 would have been sensitive enough to pick up on a *que*-effect under the true factives, had there been one. I conclude that that the absence of an effect in this case represents a genuine pattern: in non-elliptical clauses, there is no *que*-effect like in gapping. I therefore do not consider it necessary to conduct an experiment that directly compares gapping and full clauses.

As for Hypothesis 1, the data in the experiments provides evidence that the type of embedding verb has an influence on the acceptability of *que*. A comparable degradation of *que* like the aforementioned one under true factives does not exist under indicative-taking non-factives, as the results of Experiments 1 and 3 show. Like for the full clauses, I assume that the measurements in the experiments would have been sensitive enough to pick up on a *que*-effect in the indicative-taking non-factives, had there been one. These results suggest the preliminary conclusion that factivity may indeed be the determining factor in the acceptability of *que*: non-factives seem to allow *que* without restrictions, while factives show a degradation of *que*.

Experiments 5 and 6 tested semi-factives factives and subjunctive-taking non-factives. In Experiment 5, both verb types showed an effect of *que* (descriptively speaking), which only turned out to be significant for the semi-factives. This result aligns in principle with the preliminary conclusion on factivity. However, the *p*-value for the *que*-effect in the non-factives approached significance at an observed value of .05. This *p*-value, coupled with the reduced materials of Experiment 5, lead me to conduct the replication Experiment 6. This experiments had the same amount of materials as Experiments 1 to 4. The results of Experiment 6 cast doubt on the preliminary conclusion. The statistical model indicated a significant effect of *que*, and – crucially – no significant interaction effect of *que* and the type of embedding verb. This finding suggests that the effect of *que* is equally present within both verb types, thus failing to replicate the lack of a significant effect for the non-factives in Experiment 5. The results of Experiment 6 thus suggest that factivity does not determine the acceptability of *que* after all. If factivity was crucial for the acceptability of *que*, non-factives should pattern alike in not showing the *que*-effect, regardless of their mood selection. This expectation is contrary to fact: only indicative-taking non-factives consistently show no effect of *que* (Experiments 1, 3, see also Experiment 7 below for further confirmation), while subjunctive-taking non-factives show no effect in Experiment 5 but do so in Experiment 6.

The results thus far speak against Hypothesis 1. An additional issue with this hypothesis that has not been addressed yet concerns the difference between the two factive sub-types. The true factives (Experiments 2 and 3) and the semi-factives (Experiments 5 and 6) both consistently show the expected *que*-effect. But Hypothesis 1 is more detailed in regard to these two types: not only should there be an effect in both these types, the effect should be stronger for the true factives than for the semi-factives. So far, the experiments do not illuminate this aspect of Hypothesis 1. I believe a comparison of these verb types across experiments is not suited to shed light on this issue. Instead, I conducted another experiment to do so, Experiment 7.

3.3.9 Experiment 7

Experiment 7 primarily tested the three-way difference between the indicative-taking non-factives and the two factive sub-types. The comparison of these three verb types is of special interest, since these are the types that Bîlbîie & de la Fuente (2019) test in their study on locally embedded gapping. (Their study plays a major part in the inspiration of this study, see Chapters 2.3 and 2.5 for details.) This comparison will enable us to see whether the two factive sub-types pattern as expected, i.e. whether the semi-factives show a smaller *que*-effect than the true factives. The relevance of the indicative-taking non-factives in this regard will become apparent in the statistical analysis below. In addition to these three verb types, I also chose to include the fourth and last verb type, the subjunctive-taking non-factives (because why not).

The materials thus contained all 192 experimental items, which is twice as many as for any of the preceding experiments. To avoid doubling the number of items judged by each participant, the items were distributed across twice as many lists (= eight). This way, each participant judged 24 experimental items and 36 fillers, like in the preceding experiments. As a consequence of this procedure, every list contained six embedding verbs per type, i.e. half of the total embedding verbs in the materials.

As for the materials themselves, there were systematic changes to the embedded clauses. Transitive verbs replaced the intransitive and reflexive verbs. The adverbials in the antecedent and gapping clause were replaced with direct objects to the new transitive verbs. For an example of these changes, see the sample item in (57). In this example, the transitive VP *bebe vodka* 'drinks vodka' replaced the intransitive VP *fuma cada día* 'smokes every day' of the preceding experiments.[24] Some of the contexts and matrix subjects were adjusted to work better with the new embedded clauses.

(57) **Target sentence Experiment 7** (with the semi-factive verb *ignora* 'ignores')
El médico ignora que el rey beb-e vodka y {que / Ø} la reina, ron.
the doctor ignores that the king drinks-IND vodka and that the queen rum
'The doctor ignores that the kings drinks vodka and that the queen drinks rum.'

Due to the quantitative doubling of the materials, I aimed to have roughly twice as many participants as in the other experiments. In total, 108 participants completed the questionnaire. 16 of the participants

[24]The reason for this change from intransitive to transitive VPs was my initial assumption that objects show a more consistent behaviour in fronting movement than adjuncts, which would have been relevant in an analysis which involved such a movement. For some details, see the discussion on "move-and-delete" approaches to gapping and stripping in Chapter 2.2. I have since abandoned this assumption and the argument-adjunct distinction will not feature in the discussion of the data or the analysis.

were students at the Universidad de Jaén.[25] The remaining 92 participants were from Prolific.com. 48 participants were filtered out by the exclusion criteria, one of whom was a bilingual participant from Jaén. The data of the remaining 60 participants (23 female, 37 male, ages 19-50, median age 28) entered the analysis.

The results are given in Table 3.10 and Figure 3.7. Like in the preceding experiments, the indicative-taking non-factives hardly show an effect of *que*. The other three verb types each show a noticeable degrading effect of *que*, which seems to be slightly less pronounced for the semi-factives than for the other two. The subjunctive-taking non-factives and true factives show a larger effect of *que*, with hardly a difference between them.

Table 3.10: Experiment 7 (all four verb types): Mean ratings per experimental condition (*N* = number of observations, *SD* = standard deviation, *SE* = standard error, *CI* = 95% confidence interval)

Factivity	Mood	Complementizer	N	Mean rating	SD	SE	CI
Non-factive	Indicative	Ø	168	78.30	21.87	1.69	3.33
Non-factive	Indicative	*que*	171	77.64	18.94	1.45	2.86
Non-factive	Subjunctive	Ø	172	80.81	21.04	1.60	3.17
Non-factive	Subjunctive	*que*	168	70.99	22.81	1.76	3.47
Semi-factive	Indicative	Ø	169	80.24	21.47	1.65	3.26
Semi-factive	Indicative	*que*	169	74.65	22.00	1.69	3.34
True factive	Subjunctive	Ø	170	83.32	17.75	1.36	2.69
True factive	Subjunctive	*que*	166	73.56	19.38	1.50	2.97
		High acceptability controls	1003	91.55	24.65	0.78	1.53
		Low acceptability controls	997	33.89	43.68	1.38	2.71

The statistical analysis focussed on the relationship between the indicative-taking non-factives and each of the two factive sub-types. I fitted one model to each of the three possible pairings of these verb types. The three resulting models are summarized in Table 3.11. Model 1 contained the indicative-taking non-factives and the true factives. Model 2 contained the data of the two indicative-taking verb types. Model 3 contained the data of the two factive sub-types. Due to the resulting multiple comparisons, I report Bonferroni-corrected *p*-values.[26]

In all three models, the effect of the verb type is not significant (the values >1 are due to Bonferroni correction). What is more important is the question whether the verb types react differently to the presence of *que*, i.e. the interaction of the complementizer and verb type. In Model 1, which contained indicative-taking non-factives and true factives, this interaction was statistically significant, thus replicating the result of Experiment 3. Visual inspection of the bars in Figure 3.7 strongly suggests that within the indicative-taking non-factives, there is no meaningful effect of *que*, while within the true factives, there is such a meaningful difference. This impression is reminiscent of the fact that the single comparisons in Experiment 3 revealed that the *que*-effect was not statistically significant in the indicative-taking non-

[25] I thank Eugenio Olivares Merino at the Universidad de Jaén for his help in the recruitment of the participants. Jaén is a city in the north of Andalusia, Spain.

[26] NB: None of the original *p*-values in the models were such that a significant value entered the non-significant range after the Bonferroni-correction. That is, all non-significant corrected *p*-values in Table 3.11 originated from non-corrected values that were not significant either.

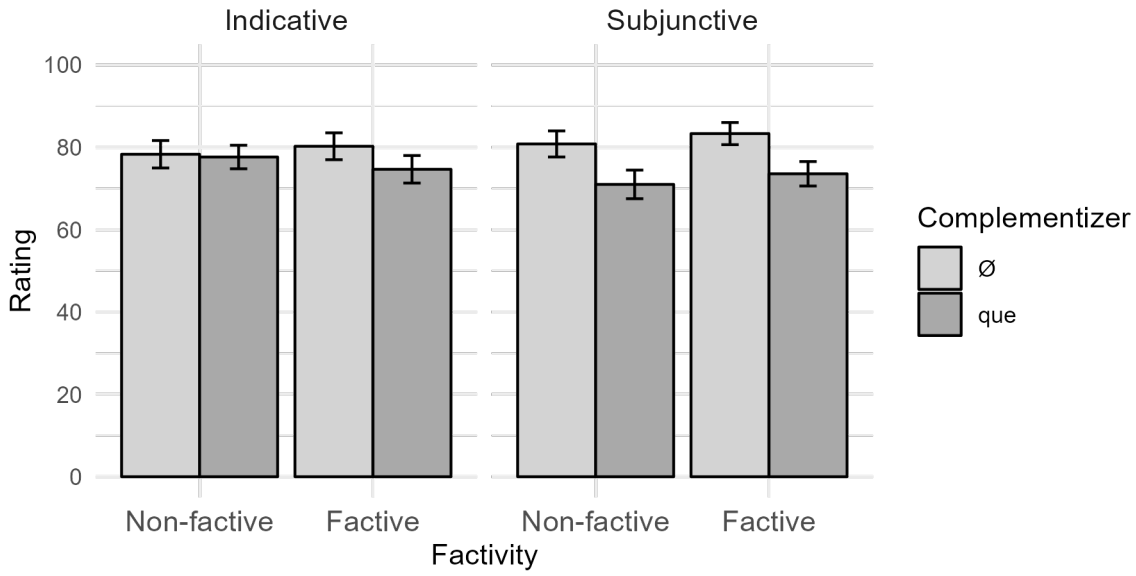


Figure 3.7: Results of Experiment 7 (all four verb types): Mean ratings with 95% confidence interval. (Filler items omitted for legibility.)

Table 3.11: The three linear mixed-effects models for Experiment 7: Experimental factors, estimates, standard deviations (SE), t-values, and p-values with Bonferroni correction for three tests.

Model	Factor	Estimate	SE	t	corrected p
1	Complementizer	2.66	0.96	2.77	0.02
(Indicative non-factives,	Verb type	0.28	0.90	0.31	>2
subjunctive true factives)	Interaction	2.20	0.66	3.32	<0.01
2	Complementizer	1.63	0.95	1.71	0.28
(Indicative non-factives,	Verb type	0.20	1.18	0.17	>2
indicative semi-factives)	Interaction	1.12	0.71	1.58	0.34
3	Complementizer	3.84	1.20	3.21	<0.01
(Indicative semi-factives,	Verb type	0.42	1.03	0.41	>2
subjunctive true factives)	Interaction	0.98	0.65	1.50	0.40

factives, but significant in the true factives. To avoid adding further instances of multiple comparisons to the analysis, I do not perform such single comparisons for the data in the present experiment.

Models 2 and 3 each contain the data from the semi-factives and compare them with the indicative-taking non-factives and the true factives, respectively. Model 2 indicates that the effect of the complementizer is not significant, thus providing no evidence that the presence of *que* affects ratings overall. The interaction of complementizer and verb type is not significant either, which suggests that indicative-taking non-factives and semi-factives do not react differently to the presence of *que* in gapping. Model 3 likewise indicates no significant interaction effect, however in this model, the effect of the complementizer is significant. These two significant values suggest that *que* equally leads to a significant degradation in semi-factives and true factives.

Models 2 and 3 thus seem to paint a contradictory picture for the semi-factives. On the one hand, they seem to pattern with the indicative-taking non-factives, which show no (statistically significant)

effect of *que* (Model 2). On the other hand, they seem to pattern with the true factives, which do show such a significant effect of *que*. I discuss this finding in detail in Section 3.4 below, where I argue that these patterns provide evidence for an intermediate position of semi-factives between (indicative-taking) non-factives and true factives, as predicted by my Hypothesis 1 (see above).

Finally, by visual inspection of the bars in Figure 3.7, the subjunctive-taking non-factives seem to show a meaningful difference between the variant with and the variant without *que*. As noted already, the size of this difference looks to be close to that of the true factives, and noticeably larger than those of the two indicative-taking verb types. On a descriptive level, this pattern seems to replicate the findings for the subjunctive-taking non-factives of Experiment 6, which showed a statistically significant effect of the complementizer. I did not perform any inferential statistics for the subjunctive-taking non-factives, in order to avoid adding more instances of multiple comparisons to the analysis.

3.4 General discussion of all experiments

The results show that the complementizer *que* is generally acceptable in embedded gapping and stripping in Spanish. This finding confirms the assumptions on gapping that I have made thus, which were based on corpus data (see Chapter 2.5). However, a closer inspection of the results indicates that while the acceptability of *que* seems to be robust in stripping, the same cannot be said for gapping. Here, I observe more intricate patterns, which is why I will focus on gapping and set aside stripping. In gapping, *que* receives degraded ratings to varying degrees depending on the type of embedding verb. Notably, however, this degradation is not so severe as to cause outright unacceptability for any of the four embedding verb types. In the non-elliptical sentences of Experiment 4 I observe no comparable effect of *que* on ratings, which suggests that the effect is not due to more general properties of clausal coordination, but instead specific to gapping.

The different degrees of degradation caused by *que* under the different verb types are visualised in Figure 3.8, which contains the collapsed gapping data from the six experiments that tested this ellipsis type (= all experiments except 4). The figure shows the differences in average ratings between the Ø- and *que*-condition for each of the four verb types. All four bars have positive values, which means that for all verb types, clauses without *que* were preferred. The preference within indicative-taking non-factives is only slight, at about 2 (out of a possible 100). All three experiments that investigated this verb type (Experiments 1, 3 and 7) showed hardly any noticeable effect of the presence of the complementizer. The statistical models of Experiments 1 and 3 confirmed this observation, as they indicated no significant effect of the complementizer on acceptability ratings. The statistical analysis of Experiment 7 contained no significance testing on the *que*-effect within the indicative-taking non-factives.

The other three verb types show a greater sensitivity to *que*, with the difference for semi-factives at about 9, for the subjunctive-taking non-factives at about 10, and the true factives at about 11. The statistical evaluations of Experiments 2, 3, and 6 indicate a significant degradation with *que* in gapping for these types. The evaluation of Experiment 5 indicated a significant effect for the semi-factives, and an effect for subjunctive-taking non-factives that was just outside the significance level at $p = .05$. However, the presence of such a significant effect in Experiment 6, which featured more extensive materials than Experiment 5, and the noticeable difference in ratings between the *que*- and Ø-conditions under this

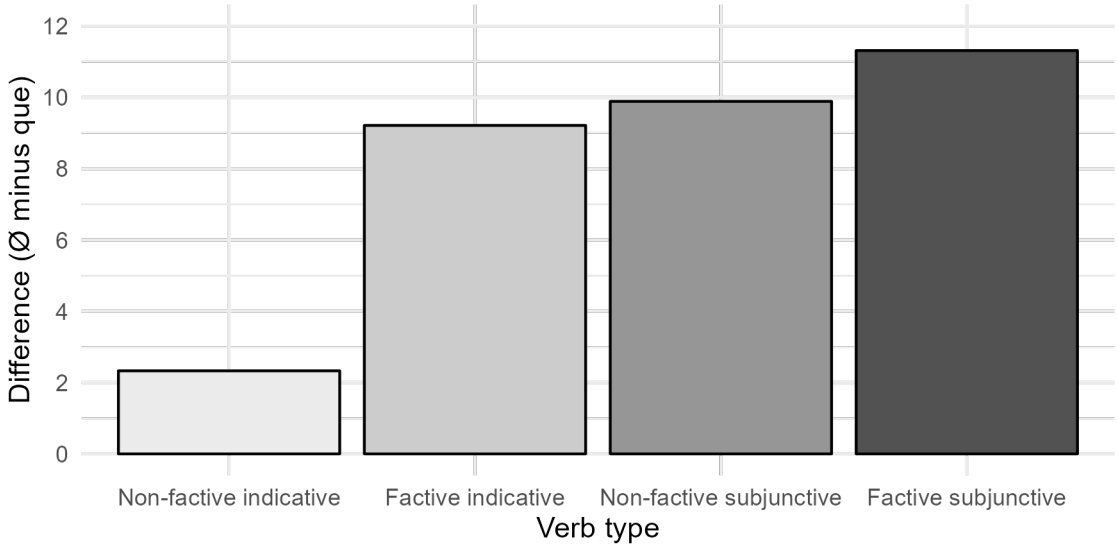


Figure 3.8: Difference of average Ø- and *que*-ratings for the four verb types (collapsed gapping data from the six experiments that tested gapping).

verb type in Experiment 7 leads me to conclude that the degradation in ratings with *que* observed in the experiments represents a genuine pattern for these verbs. Like for the indicative-taking non-factives, the statistical evaluation of Experiment 7 did not perform significance tests on the *que*-effect within the subjunctive-taking non-factives.

These results lend some credence to the Typological Conjecture of Chapter 2, reproduced here in (58). Descriptively speaking, the three verb types examined in both Bîlbîie & de la Fuente's (2019) study on local embedding, and the present study on global embedding are in the same order as regards the measured effect: (indicative-taking) non-factives show the smallest effect, true factives show the largest effect, and semi-factives range in between.

(58) **Typological Conjecture on Embedded Gapping**
Any constraint on a complementizer specific to ellipsis that operates in local embedding also operates in global embedding (and vice versa).

The statistical evaluation of Experiment 7 supports this description. The linear model containing the indicative-taking non-factives and the true factives returned a significant interaction of verb type and *que*, which suggests that *que* affects ratings with the factives differently than with the non-factives. However, neither of the other two linear models fitted to the data of Experiment 7, which compare the semi-factives to the indicative-taking non-factives, and to the true factives, respectively, return significant interaction effects. I interpret this lack of interaction effects as evidence that semi-factives simultaneously pattern with both the indicative-taking non-factives, and with the true factives, even though these two types pattern differently from each other. I propose that these statistical results may be interpreted as evidence for the intermediate position of the semi-factives in between the indicative-taking non-factives and the true factives. This interpretation is congruent with the conclusion I have drawn from the descriptive relative order of the types in Figure 3.8.

The last of the four verb types in my experiments are the subjunctive-taking non-factives. In Figure 3.8, this verb type shows a noticeable *que*-effect, which ranges in between the two factive types, with a noticeable difference to the indicative-taking non-factives. This pattern is unexpected under the assumption that factivity determines the acceptability of *que*. My experimental findings thus do not provide evidence for the hypothesis that factivity is the decisive factor for the acceptability of *que*, since the two groups of non-factives behave differently. A similar reason rules out mood selection as the decisive factor: while both subjunctive-taking verb types display a significant penalty on *que*, only one of the indicative-taking ones does so (i.e. the semi-factives). Mood selection is no reliable predictor for the acceptability of *que* either.

The presence of an effect of *que* in subjunctive-taking non-factives and both types of factives suggests the descriptive generalization that these three verb types form a joint class (with, potentially, variation within the class), to the exclusion of indicative-taking non-factives. However, a close look at the individual verbs reveals that neither this two-way split, nor the split into the four verb classes neatly captures the variation of the acceptability of *que*. Figure 3.9 displays the difference between Ø- and *que*-ratings for every individual embedding verb. The figure shows that the verbs of none of the four types exhibit a strictly uniform pattern. There are no clear-cut borders between verb types, and differences between types only surface as tendencies. For instance, as expected from the average differences between verb types, indicative-taking non-factives tend occur towards the lower end of the scale. Two of these verbs (*considera* 'considers', 'perceives' and *imagina* 'imagines') even show a negative value, which means that on average, the items they ocurred in received higher ratings with *que* than without *que*. The other verb that shows a preference for *que* (albeit to a very small degree) is the semi-factive *acepta* 'accepts'. Conversely, subjunctive-taking true factives tend to occur towards the higher end of the scale. However, some indicative non-factives incur a relatively large *que*-penalty (e.g. *relata* 'relates'), and some subjunctive factives incur a relatively small one (e.g. *lamenta* 'regrets' and *critica* 'criticizes'). The verbs of the other two types display a similarly variable behaviour.

I see two options on how to deal with these findings in terms of factivity. First, I may actually hold on to the hypothesis that it is factivity that determines the acceptability of *que*. This hypothesis may potentially be coherent with my experimental results, if I assume that the categorization of the individual verbs as factive or non-factive was incorrect in many cases. This assumption opens up the possibility that those verbs showing a large degrading effect of *que* are factive after all, and those that show little or no degradation are non-factive after all. This scenario would thus save the hypothesized factivity-*que* dependency, and is actually not unlikely given that the categorization of the verbs proceeded on the basis of my intuitive judgments on a language I do not speak (see Section 3.1.1). In Chapter 4 I amend this empirical shortcoming by providing quantitative, native speaker data for the clause embedding verbs in my experiments. In that chapter I adopt a gradual view of factivity in the hopes that it provides a point of connection to the gradualness of the *que*-effect I observed here. The data in that chapter shows that my categorization of the verbs was by and large accurate, which like the results of the experiments suggests that factivity does not determine the acceptability of *que*.

As a second option, I may reject the hypothesis that it is factivity specifically that determines the acceptability of *que*. Doing so would be the obvious way to go since it is coherent with my experimental results without calling into question the fundamental assumptions as to the categorization of the verbs.

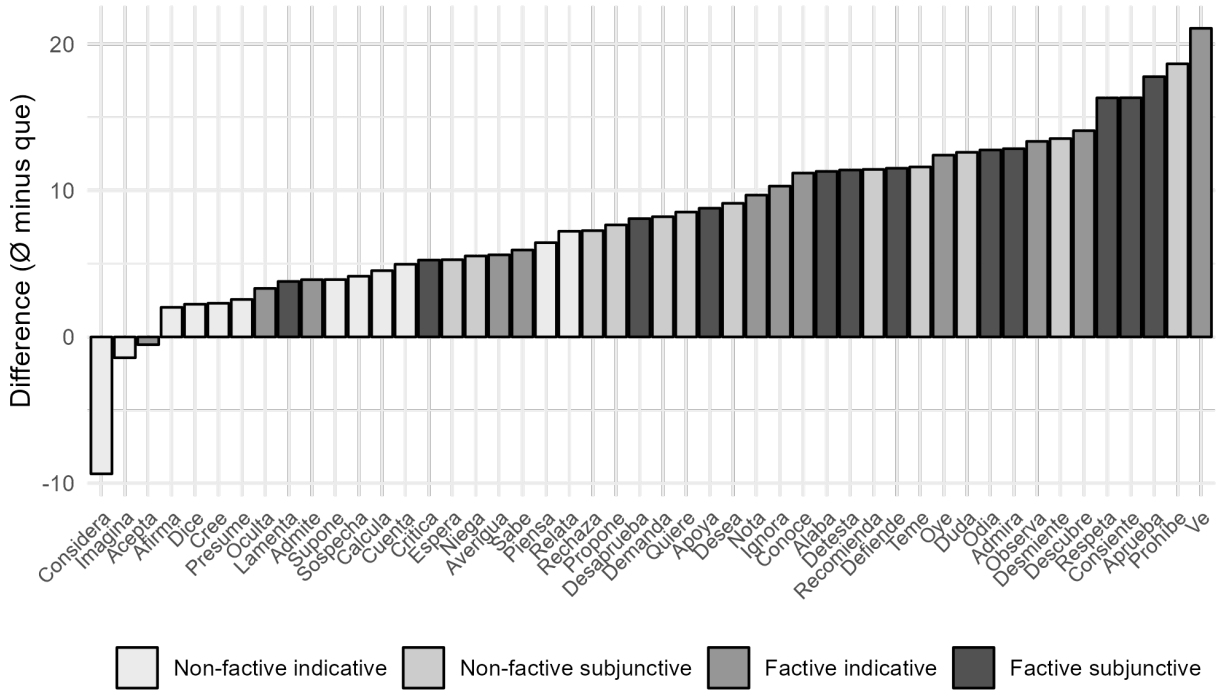


Figure 3.9: Difference of average Ø- and *que*-ratings for each verb (collapsed gapping data from the six experiments that tested gapping).

But then the question arises of what determines acceptability instead. One obvious hypothesis would be that instead of factivity, a different meaning component of the verbs is the determining factor. This component should characterize the indicative-taking non-factives, since these are the only verb type that consistently shows a negligible effect of *que*. I go into more detail on this option in the following section, and explore it experimentally in Chapter 5.

The following section also deals with the question of how *que* might interact with the meaning of the embedding verb such that the observed degrading effect on acceptability arises in the first place. After all, there is no obvious connection of *que* to factivity since it does not seem to carry any meaning related to factivity, or to any other relevant verbal property for that matter. All embedding verbs select for this complementizer, whether they are factive or not. An additional complication of this issue arises when we consider that *que* only affects acceptability in gapping, but not in full clauses. This observation is unexpected if we assume that it is a meaning component of *que* itself that triggers the effect, which predicts the same effect in elliptical and full clauses. In the following section I outline a possible answer to this question. I argue that *que* in gapping is indicative of a marked syntactic structure, which is responsible for its effect with some embedding verbs.

3.5 Sketching a structure for gapping with *que*

The presence of the complementizer *que* in embedded gapping in Spanish has not played a central role in any in-depth theoretic analysis of this type of ellipsis. This is despite the fact that embedded environments

may provide important insights for the long-standing discussion of what syntactic objects are coordinated in gapping: 'large', i.e. clausal, or 'small', i.e. sub-clausal conjuncts (see Chapter 2.2 for details on this question and a number of existing approaches for English). In this section, I provide a rough sketch on how existing proposals from the literature on gapping on the one hand, and on the structure and interpretation of complement clauses on the other hand could work together to go some way towards an explanation of the data in my experiments. I will adopt a 'hybrid' approach to gapping, which involves both clausal and sub-clausal coordination for embedded gapping with and without *que*, respectively. Again, see Chapter 2.2 for a description of some of the different approaches in English. For 'hybrid' approaches to Spanish gapping, see Centeno (2011) and Jung (2017, 2023).[27] Since gapping with *que* showed a sensitivity to the type of embedding verb, the relevant structure (i.e. clausal-coordination gapping) will then be scrutinized more closely. I outline how the truncation of functional structure in (some) complement clauses might account for the degradation in the data, and how this truncation might intertwine with effects on interpretation. For ease of exposition, I will assume that the categorization of the verbs as factive or non-factive in my experiments is correct, and that therefore factivity as such is not responsible for the degradation.

In order to arrive at a structural analysis for gapping, I will take a short detour via stripping. Recall that in my data, stripping and gapping share the characteristic that *que* is in general optional, the difference being that gapping shows degradation for the structure with *que* if embedded under certain verbs. Since embedded stripping has received more attention in the literature than gapping (see the introduction to this chapter), I will use stripping as a starting point for my discussion to explain the general optionality of *que* and then turn to the degradation in gapping. Concretely, I will capitalize on Fernández-Sánchez's (2019) proposal on stripping, which accounts for the optionality of *que* by a 'hybrid' analysis involving variably clausal or sub-clausal coordination. Sub-clausal conjuncts do not include a structural position for the complementizer (i.e. the head of CP), and therefore no *que* surfaces in the second conjunct, see (59). Clausal conjuncts, on the other hand, include such a position for *que*, which is why it surfaces, see (60).[28]

(59) [*embedding verb* [[$_{CP}$ *que* [[$_{XP}$…] & [$_{XP}$ …]]]]] (sub-clausal coordination)

(60) [*embedding verb* [[$_{CP}$ *que* …] & [$_{CP}$ *que* …]]] (clausal coordination)

An advantage of this proposal is that the optionality of the complementizer is entirely structural and thus independent of the ellipsis itself, thereby mirroring the optionality of the complementizer in the second of two coordinated, non-elliptical complement clauses. I observe such optionality regardless of the type of embedding verb in Experiment 4. No additional ellipsis-specific rule needs to be stipulated, which would be necessary in invariant (sub-)clausal conjunct approaches. I will adopt this proposal for embedded gapping.

[27] I do not discuss the contributions by Centeno (2011) and Jung (2017, 2023) in depth since they do not feature clausal embedding.

[28] Fernández-Sánchez (2019) identifies the relevant coordinated phrases as CPs for clausal coordination and TP/F(ocus)P for asymmetric sub-clausal coordination. I follow Fernández-Sánchez (2019) in assuming that clausal coordination occurs at CP, but I will remain non-committal as to the exact identity of the sub-clausal constituents (TPs, *v*Ps, VPs etc.). I will also remain non-committal to the exact mechanism governing ellipsis (move-and-delete, ATB-movement, *in situ*-deletion etc.), since it is immaterial to the discussion at hand.

The 'hybrid' approach predicts that gapping with *que* and gapping without *que* are both grammatical. As such, it covers a major part of my experimental findings, in that *que* was judged overall acceptable for all verb types in all experiments. However, the approach is not yet fit to account for the more intricate patterns I found concerning the degradation of *que* under some of the verbs. To approach an explanation of these intricate patterns, I will take a closer look at the structure of the elliptical conjunct in clausal conjunct gapping (i.e. gapping with *que*). Since gapping without *que* in comparison shows a uniform pattern in the data, I will not scrutinize gapping with sub-clausal conjuncts in the same way.

I will follow the popular approach in the literature that remnant constituents occupy left-peripheral positions in the clause. This has been adopted for Spanish gapping (Centeno 2011), as well as for other types of ellipsis (such as stripping, Fernández-Sánchez 2019, Villa-García 2016).[29] According to this approach, the remnant constituents do not occupy their thematic positions lower down the clause, but instead occupy positions in the CP-domain (i.e., outside the TP).[30] I will return to the structural details below, because some further aspects of the structure of embedded clauses need to be addressed first. What is relevant for now is that I can speculate that, rather than gapping itself, it is these left-peripheral positions that are responsible for the variable degradation of *que* under different verbs. To outline this speculation, I will turn to the internal structure of embedded CPs, and the way this structure interacts with the interpretation of the clause.

As put forward by Haegeman (2006), the structure of an embedded clause in general depends on the type of embedding verb. Specifically, Haegeman (2006) observes that in English, a complement clause may not host a left-dislocated element, if the clause is embedded under the factive verb *regret*. She therefore concludes that complements of factives have a truncated left periphery, which lack the structural positions for dislocated elements. These are the same positions that the remnant constituents occupy in clausal conjunct gapping. I have reason to assume that the same truncation is at play in Spanish complement CPs[31] because of independent evidence from recomplementation, i.e. the occurrence of multiple complementizers in the same embedded clause. Example (61) illustrates that a clausal constituent that is left-dislocated in an embedded clause such that it directly follows the complementizer *que*, may itself be followed by a second *que*. In (61), the dislocated adverbial PP headed by *a* 'at' is sandwiched between two *que*s. Villa-García (2016) analyses the second *que* as the head of a left-peripheral phrase, with the dislocated element occupying the specifier of the phrase. For a rough structural sketch, see (63).[32] Under factive verbs, dislocation only seems to be available without the second *que* as in (62), where the dislocated DP *ese coche* 'that car' cannot be followed by a second *que* because its host CP is embedded under factive *lamentar* 'regret' (Demonte & Fernández-Soriano 2009). I take the obligatory absence of secondary

[29] For similar approaches to English, see Chapter 2.2.

[30] These positions are usually thought to fulfill information structural functions such as Focus or Topic. My proposal is contingent with this, but I will remain non-committal as to the exact identities of the phrases in question, since I am not concerned with issues of information structure here. For similar reasons, I will not discuss the mechanism by which the remnants arrive in these positions (i.e. movement or base generation).

[31] See also de Cuba & MacDonald (2013) and García Núñez (2020), who assume similar truncation approaches for Spanish.

[32] A common way of specifying the left-peripheral phrases in question is by means of Rizzi's (1997) cartographic approach, where the phrases hosting primary and secondary *que* are identified as ForceP and FinP, respectively, and the position hosting the dislocated XP as TopP (e.g., de Cuba & MacDonald 2013, Demonte & Fernández-Soriano 2009, Villa-García 2016). I will not discuss this, see footnote 30. For illustrative purposes, I use the more neutral placeholder labels *CP1* and *CP2* in (63).

que to indicate that in factive complement clauses, left-peripheral positions are indeed not available for dislocated elements, see (64). These must occupy positions lower in the clause.

(61) Dice **que** a pesar del fracaso electoral, (**que**) nadie ha dimitido.
says that at despite of+the fail electoral that nobody has resigned
'S/he says that despite the electoral disaster, nobody resigned.' (Villa-García 2019: 3)

(62) Lamento **que** ese coche (***que**) no lo compres.
am.sorry that that car that not CL.3SG buy-2SG.SUBJ
'I am sorry that you won't buy that car.' (Demonte & Fernández-Soriano 2009: 47)

(63) [$_{CP1}$ *que* [$_{CP2}$ [dislocated XP] (*que*) [$_{TP}$ …]]] (non-factive complement CP)

(64) [$_{CP}$ *que* [$_{TP}$ …]] (factive complement CP)

To summarize: I assume that non-factive complement clauses have a rich left periphery that includes positions for dislocated phrases, and that factive complement clauses lack these positions. As I discussed above, I also assume that in clausal coordination gapping, remnants must occupy left peripheral positions. As per these assumptions, clausal conjunct gapping is barred under factives because the complement clauses of factives do not have the relevant left-peripheral positions to host the remnant constituents. A *que* in the gapping conjunct thus not only indicates that the conjunct is a CP, but also that it is a structure you would expect for the complement clause of a non-factive verb. With factives, *que* would thus indicate a mismatch between the type of the embedding verb and the structural make-up of the embedded clause. As a result, acceptability ratings of such structures would be expected to drop. This proposal predicts sensitivity to verb type as is required to account for the data from my experiments. However, it does not make quite the right predictions. First, the violation of a syntactic constraint should result in outright unacceptability rather than the gradual degradation I observe in the data.[33] Second, the approach makes use of the non-factive/factive distinction, which likely is not the decisive characteristic for the acceptability of *que* (see Section 3.4). Rather, I observed that degradation is present largely, but not only, under factives.

To resolve this tension, let me return to Haegeman's (2006) discussion of dislocation in English. Recall that, similar to what I assume for *que* in embedded gapping, dislocation in complement clauses is thought to be indicative of a non-factive structure. As Haegeman (2006) notes, if the complement CP of *regret* (a factive verb) hosts a dislocated element, such as the adjective phrase headed by *due* in (65), the sentence does not become outright ungrammatical. Instead, the interpretation of the sentence shifts: It no longer contains "the subject's emotive reaction to a particular 'fact'" (Haegeman 2006: 1666), but an embedded assertion, i.e. an utterance with which new information is introduced into discourse. It thus differs from a factive clause, which contains a proposition whose truth is presupposed.

(65) I regret that due to a funding shortage there will no longer be any drinks available at the bar for non members. (Haegeman 2006: 1666)

[33] Bîlbîie & de la Fuente (2019), who observe gradual effects on acceptability in locally embedded gapping (Chapter 2.5), reject the possibility of an approach involving two distinct structures for complement clauses for precisely this reason: two distinct structures predict two distinct judgment patterns (acceptable vs. unacceptable), which they argue is not suitable to explain gradual differences. I discuss how to arrive at the necessary gradualness under the assumption of exactly two structures in the following.

For embedded gapping with *que*, I will assume that the complementizer, like the dislocated element in (65), signals that the embedded clause is connected to assertion (rather than 'non-factivity' as such), because it is part of a rich left periphery under clausal coordination.[34] Importantly, I assume that a gapped clause introduced by *que* can embed under any verb without creating a syntactic mismatch (and thus without causing outright ungrammaticality). However, depending on the verb there may be a semantic/pragmatic mismatch that leads to difficulties in interpretation. For certain verbs, such as verbs of saying and thinking (= the indicative-taking non-factives in my experiments), there will be no issue, since these can straightforwardly embed assertions and would thus not receive degraded ratings. For other verbs including factives and non-assertive non-factives (= the other three types in my experiments), the required shift in interpretation will lead to a degradation in acceptability. Assertive readings under these verbs will be more difficult for participants to coerce, or result in inappropriate sentence meaning. This, in turn, will bring down acceptability judgments somewhat, but not enough to indicate ungrammaticality. These assumptions account for the degradation of *que* I observe in the verb types in my experiments: Many verbs that easily embed assertions are indicative taking non-factives, which as a group display the smallest degradation for *que*. The other three types largely include verbs that are less suited to embed assertions (i.e. factives and non-assertive non-factives) and as groups incur notably larger penalties. However, the required shift towards an embedded assertion has a different degradation size effect on the individual verbs, leading to the gradualness I observe across verbs in my data.

To illustrate how such a shift in interpretation might proceed, let us look at a specific example from my experimental materials. The sample item from Experiment 3 in (51) above is reproduced in its version with *que* in (66). The embedding verb is *desaprueba* 'disapproves', which I classed as factive in the experiments. If my analysis is on the right track, I should arrive at a degraded (but not unacceptable) rating roughly in the following way: The presence of *que* in the second conjunct indicates that the conjuncts must be clausal. In the context of gapping, the structure involves the left-peripheral positions for the remnants, which means that the complement clause has a non-truncated, 'large' structure (see 63 above). This large structure is that of an assertive clause, which is incongruous with the 'normal' factive reading of the embedding verb. However, I can 'rescue' the sentence by shifting the expected or commonly used factive reading of the embedded clause to an assertive reading. The resulting interpretation of (66) is thus not *Santiago disapproves of the fact that* p (where *p* = the CC), but rather an assertion that *p* and that *Santiago has a disapproving attitude towards* p.

(66) Target sentence Experiment 3, factive conditions
Santiago desaprueba que el rey fume cada día y {que / Ø} la reina, a veces.
Santiago disapproves that the king smokes-SUBJ every day and that the queen at times
'Santiago disapproves of the fact that the king smokes every day and that the queen smokes sometimes.'

As I mention above, an advantage of my proposal is that it capitalises on general structural properties of embedded clauses and is not restricted to globally embedded gapping. Thus, there should in principle be similar structures to which it could be applied. Locally embedded gapping is a good candidate to be

[34] Also recall that I assume that the remnant constituents in embedded gapping with *que* are dislocated. They occupy the same positions as the dislocate in (65). However, since in gapping the exact position of the remnants (left-peripheral vs. thematic) is not transparent, the remnants themselves are not indicative of the structure of the clause.

one of those structures: The constituent containing the gap must be a CP, because it is selected for by a clause-embedding predicate. It is thus (invariably) structurally equivalent to the large (= clause-sized) conjunct I propose for embedded gapping with *que*, including the dislocated positions for the remnants. The assertive interpretation associated with this syntactic object is exactly what is required in locally embedded gapping, since the elliptical clause must receive the same assertive interpretation as the non-embedded assertive clause it is coordinated with (i.e. its antecedent).

To illustrate, consider the sample item from Bîlbîie & de la Fuente (2019) experimental materials in (67). Recall that in one of Bîlbîie & de la Fuente's (2019) experiments, items were presented in the context of an anchor question like (67a). Since both the antecedent and the embedded elliptical clause in their test sentences (see 67b) provide answers to the anchor question, they must be interpreted as assertions (cp. the discussion of assertion in Section 3.1.1). Under this view of assertion embedding, I can derive the sensitivity to the type of embedding verb in locally embedded gapping the same way like in globally embedded gapping: the difficulty with which an assertive reading is obtained under embedding verbs that usually do not embed assertions causes a degradation in ratings. This applies for example to ordinarily factive verbs like *molesta* 'bothers' in (67b). However, since an assertive reading (as per assumption) is not impossible to obtain for these verbs, ratings are not so low as to indicate outright unacceptability. Conversely, for verbs that easily embed assertions (= non-factive verbs of saying and thinking), there will be no degradation. This analysis of locally embedded gapping differs only minimally from that of Bîlbîie & de la Fuente (2019), who also make assertion embedding a central part to their analysis. However, I reckon my view is still superior because it avoids some of Bîlbîie & de la Fuente's (2019) conceptual shortcomings regarding the interplay of factivity and assertion (see Chapter 2.5). Furthermore, my analysis provides a way to generalize the effects to both configurations of embedded gapping, and does not rely on mismatches between the conjuncts like Bîlbîie & de la Fuente's (2019) analysis.

(67) a. ¿Qué bebidas pidieron los dos amigos en el bar?
what drinks ordered the two friends in the bar
'What drinks did the two friends order in the bar?'

b. Pablo pidió una cerveza y me molesta que Juan un whisky.
Pablo ordered a beer and me bothers that Juan a whisky
'Pablo ordered a beer and I am bothered that Juan ordered a whisky.'

(from Bîlbîie & de la Fuente's 2019 materials, see Chapter 2.5.2)

Another fairly immediate parallel can be drawn to embedded polarity answers. As argued for by de Cuba & MacDonald (2013), the inability of some verbs (including factives) to embed polarity answers, e.g. *lamento* 'regret' in (68), may be due to a lack of functional structure in the (in this case truncated) complement clause. However, corpus examples of embedded polarity answers under *lamentar* like (69) can be found. As a matter of fact, my informants tell me that *lamenta* is interpreted as a verb of saying in (69), which my translation reflects. Thus, I may hypothesize that the larger functional structure in the complement clause causes an interpretative shift from factive to assertive, similar to what I hypothesize about embedded gapping with *que*.

(68) A: Did they arrive on time?

B: Creo / *Lamento que sí / no.
think regret that yes no.
'I think/regret so/not.' (de Cuba & MacDonald 2013: 312)

(69) [Context: In a survey, participants were asked "whether the lack of governance was detrimental to the economy".]
Un contundente 71,7% lamenta que sí.
a overwhelming 71.7% regrets that yes
'An overwhelming 71.7% regrettingly say yes.' (from a Spanish online newspaper)[35]

Another structure in which the presence of *que* affects the interpretation of the complement clause, involves embedded questions. In Spanish, one way of introducing an embedded question is like in English by a 'bare' *wh*-pronoun or interrogative complementizer. Alternatively, unlike in English, the non-interrogative complementizer *que* may precede the *wh*-pronoun or interrogative complementizer (Rivero 1978, 1980, Plann 1982, Suñer 1993, 1999, Lahiri 2002, Demonte & Fernández-Soriano 2009, García Núñez 2020). The latter variant is in (70), where *que* precedes the *wh*-pronoun *cuándo* 'when'. In these structures, *que* shows a similar kind of sensitivity to the type of embedding verb: it is acceptable with verbs of saying like *pregunta* 'asks' in (70), but it is not acceptable with verbs of knowing like *supieron* 'knew' in (71).

(70) María pregunta **que** cuándo habéis comprado los regalos.
María asks that when have.2PL bought the presents
'María asks when you have bought the presents.' (García Núñez 2020: 56)

(71) Ya supieron (***que**) por qué lo habias hecho.
already knew-3PL that for what it had-2SG done
'They already knew why you had done it.' (adapted from Plann 1982: 300)

The existence of verb type constraints is not the only point of similarity between embedded gapping and embedded questions with *que*. There are also syntactic parallels: under *que*, the clause has a greater left peripheral structure than the equivalent clause without *que* (García Núñez 2020). This is evident in clauses involving left dislocation: When a dislocated constituent precedes the *wh*-word, *que* obligatorily surfaces in front of that dislocated constituent. This pattern is illustrated in (72), where the dislocated direct object *los regalos* 'the presents' is sandwiched between *que* and the *wh*-word *cuándo* 'when'.

(72) María pregunta ***(que**) los regalos cuándo los habéis comprado.
María asks that the presents when CL.3PL have.2PL bought
'María asks when you have bought the presents.' (García Núñez 2020: 56)

These similar properties regarding verb type constraints and a rich left peripheral clause structure suggest that the *que* in embedded questions may be the same *que* we find in the other structures I have discussed so far. Obviously, my proposal involving assertion cannot be transferred to embedded questions one to one, since embedded questions like those in (70) and (72) can hardly receive an assertive interpretation. I would like to speculate that one way to remedy this problem would be to propose an analysis in terms of at-issueness instead of assertion. At-issueness may be a good way for expanding the assertion-centric proposal of the structures I have looked at so far, which all involve declarative clauses (i.e. both types of embedded gapping, recomplementation, and fragment answers) to questions. This means that a rich left periphery in embedded clauses causes an at-issue reading rather than an assertive reading specifically. For embedded declarative clauses, an at-issue interpretation of the CC will coincide with an assertive

[35] https://www.elmundo.es/espana/2016/02/21/56ca224046163f3b3d8b45e6.html, slightly adapted. Accessed 24 Sept. 2019.

interpretation, which preserves this part of my analysis for the declarative structures. For embedded questions on the other hand, an at-issue interpretation will promote the content of the embedded question to speech-act level. That is, with *que*, an embedded question will more likely be interpreted as a request to the interlocutor (like a non-embedded question) rather than just the report of a question someone else has made. (However, this reportive meaning component will likely not be entirely lost.) As I said, this idea is speculative, more research into this specific facet of the interpretation of embedded questions is needed. The central point to this discussion is that at-issueness may, next to or in addition to assertion, be a candidate for the relevant meaning component to account for the verb-type effects on *que* in the structures in question.

The existing research on the interpretation of *que* in embedded questions makes a different observation. According to one relatively wide-spread proposal, *que* marks an indirect quotation (Plann 1982, Lahiri 2002). Under this view, *que* in embedded questions thus functions as a reportive (or quotative) marker. The pair of examples in (73) shows what this means: a sentence that features an embedded question introduced by *que* like (73a) can be paraphrased as a sentence where the embedded clause is replaced with an independent clause that contains direct speech, see (73b).

(73) a. El psiquiatra nos {dijo / repitió } muchas veces que por qué lo habríamos hecho.
the psychiatrist us told repeated many times that for what it could.have-1PL done
'The psychiatrist asked us many times why we could have done it.'

b. El psiquiatra nos {dijo / repitió } muchas veces, "¿Por que lo habreis hecho?"
the psychiatrist us told repeated many times for what it could.have-2PL done
'The psychiatrist said/repeated to us many times, "Why could you have done it?" '

(Plann 1982: 301)

As discussed in detail by García Núñez (2020), this interchangeability of embedded questions with *que* and direct quotation holds in many, though not all, cases. For this reason, García Núñez (2020) follows a slightly different approach in terms of speech-act embedding rather than reportiveness. The more intricate details of this difference between reportiveness/quotativeness and speech-act embedding is not relevant here. The relevant point to this discussion is that like at-issueness above, the special meaning component that *que* contributes to embedded questions, in tandem with a rich left-peripheral structure, may contribute to a potential explanation of the verb type effects in these structures. I will continue this thought under the assumption that *que* has a reportive function in embedded questions and, by extension, in the other structures I have discussed.

The basic reasoning to account for the verb type effects in globally embedded gapping with reportiveness is the same as for assertion and at-issueness: verbs of saying and thinking naturally carry a reportive meaning, thus causing no degradation when they embed a structure headed by reportive *que*. Other verbs (i.e. the other three verb types in my experiments) do not easily allow for such a reportive reading, thus causing a degradation in acceptability in the same structural environments, when they are coerced into this meaning.

Support for the claim that *que* may trigger a reportive interpretation comes from its use in main clauses. Main clauses are a relevant environment for this discussion, since they feature a full clausal structure, including positions for dislocated elements. When *que* introduces a main clause like in (74), it gets a

reportive interpretation.[36] Reportive *que* thus seems to be a general feature of Spanish grammar (Etxepare 2010, Corr 2018, 2022), which reinforces the hypothesis that the *que* in globally embedded gapping and the other structures discussed in this section might be this reportive *que*.

(74) Era el becario… **Que** le ha tocado la lotería.
was the intern that to.him has touched the lottery
'It was the intern... [He said] he's won the lottery.' (naturally occurring example from Corr 2018, slightly adapted)

The hypothesis that reportiveness is crucial in this discussion finds further support in the meaning shift of factives we observed in false belief context. Recall from the discussion in Section 3.1.1 that in contexts involving false beliefs, the meaning of the CC of a factive embedding predicate may change from its most natural, factive meaning to that of a "free indirect speech" (Abrusán 2011). It is likely that what Abrusán (2011) mean by "free indirect speech" is in fact a reportive meaning. Recall also that this meaning shift not only obtains in contexts which made the false belief explicit, but may also be triggered by other means. (Section 3.1.1 briefly discusses prosody in this regard.) I may thus speculate that the shift towards a reportive reading is not specifically tied to false belief contexts, but can surface by grammatical means. Specifically, the marked syntactic structure involving *que* and a rich left periphery discussed in this section represents a grammatical structure that triggers a reportive meaning.

To sum up this section, I combined existing approaches to gapping (and stripping) on the one hand, and the structural truncation of embedded clauses on the other hand, to sketch an analysis of globally embedded gapping with *que*. I showed how the gapping-specific interaction effects of *que* and the types of embedding verbs that I observe in my experimental data may arise. I did this under the assumption that my categorization of the embedding verbs as factive and non-factive has been correct, which meant that factivity cannot be the trigger for these effects. I proposed three possible alternatives which more closely characterize the verb type that shows no *que*-effect, i.e. the indicative-taking non-factive verbs of saying and thinking. These alternatives are assertion, at-issueness and reportiveness, all three of which enable us to better account for the patterns of *que* in globally embedded gapping, and to locate these patterns within a more general picture with further structures that also show similar verb type effects. The discussion around these alternative meaning component was largely speculative. I will scrutinize these three properties more closely as to their effect on *que* in globally embedded gapping in Chapter 5.

A final caveat to this unifying aspect of the analysis comes from my own data. Since I transferred the structure I assume for clausal-coordination gapping from Fernández-Sánchez's (2019) and Villa-García's (2016) treatments of stripping, the two types of ellipsis should be no different when it comes to *que*. However, my Experiments 1 and 2 do indicate differences. Stripping does not undergo the same degradation with *que* with factives that gapping does. I acknowledge this problem, but will not attempt to solve it here. More empirical work is needed on the differences between gapping and stripping in embedded environments.

[36] The reportive reading with main clause *que* is unlikely due to the ellipsis of an embedding verb of saying, since the context need not feature such a verb for *que* to have a reportive interpretation. This is the case in (74), where the context features no possible antecedent to an elided embedding verb. It thus must be *que* itself that carries this reportive meaning, which it does, according to my assumption, in combination with a rich left peripheral structure. Due to the lack of dislocation in (74), his structure is not visible in this example.

3.6 Conclusion

This chapter contained an experimental study on the complementizer *que* 'that' in Spanish globally embedded gapping. The central result of this study is that *que* is overall acceptable, which goes against introspective judgments in the theoretic literature (Fernández-Sánchez 2021). This result thus provides novel and robust data on this aspect of ellipsis. Overall, this general result represents the most substantial contributions of this chapter (and of the dissertation as a whole) to Spanish linguistics and ellipsis typology.

The largest part of this chapter, however, was concerned with more nuanced hypotheses. Specifically, I hypothesized that *que* displays restrictions on the type of embedding verb that are comparable to those working on locally embedded gapping (Bîlbîie & de la Fuente 2019). To test this hypothesis, I conducted seven experiments on the acceptability of *que* globally embedded gapping. My results indicate that *que* may generally surface in the elliptical conjunct, but depending on the (type of) embedding verb, *que* causes varying degrees of degradation. There is no meaningful degradation under (indicative-taking) non-factives, and a noticeable degradation under true factives. This degradation is specific to gapping, as in full clauses, *que* does not lead to such a degradation with true factives. Similar to the true factives, semi-factives also show a degrading effect with *que*, which ranges in size between that of the (indicative-taking) non-factives and the true factives. My results thus parallel those of Bîlbîie & de la Fuente's (2019) on locally embedded gapping. These similar results lend credence to the Typological Conjecture, which hypothesizes that in any individual language locally and globally embedded gapping are subject to the same the same constraints (Chapter 2). However, the fourth verb type I examined, subjunctive-taking non-factives, as well as the ratings for the individual embedding verbs outside of this four-way categorization suggest that factivity itself is unlikely to be entirely responsible for the effect of *que* on acceptability.

I nevertheless held on to the assumption that a meaning component of the embedding verbs is responsible for the degradation. With this assumption, I sketched an account of the sensitivity of *que* to the verb type in gapping, which was based on the syntactic ambiguity of gapping and the structural truncation of complement clauses. The degradation, I argued, stems from the interplay of clausal structure and interpretation: The easier the interpretation of the relevant meaning component is obtained for a given verb, the smaller the *que*-penalty will be. This analysis may be unified with existing accounts of other further structures that display similar sensitivity to the type of embedding verb, including locally embedded gapping, embedded fragments, embedded questions, and potentially further structures.

If the *que*-effect is indeed driven by a meaning component of the embedded verbs, of course the question arises what meaning component is it. I outlined two general possibilities: first, my assumptions on the category membership of the embedding verbs (non-factive vs. factive) may have been incorrect in some (or even many) cases. If I miscategorized enough of those verbs that show a noticeable degradation with *que* as non-factive, and enough of those that show a smaller degradation as factives, the relevant meaning component might be factivity after all. Second, another meaning component may be relevant for the mismatch. As I speculated in my structural account, this meaning component may relate to assertion, at-issueness or reportiveness. All three of these properties are possibly relevant since they characterize the indicative-taking non-factives (verbs of saying and thinking), which in marked contrast to the other three types show a negligible *que*-penalty in globally embedded gapping.

In the following two chapters, I discuss these two options in turn. Chapter 4 investigates whether my categorization of the verbs was accurate, and whether a more fine-grained approach to factivity allowing for gradualness might explain the gradual *que*-effect across verbs that I observed in this study. Chapter 5 explores whether the complementizer triggers a difference in interpretation as regards factivity, assertion embedding, at-issueness or reportiveness. To foreshadow the results of these two chapters, while my categorization of the verbs was largely accurate, there seems to be no effect of *que* on any of the interpretive properties.

4 An experimental study on the factivity of Spanish clause-embedding verbs

Presuppositions have long been discussed on the basis of largely introspective judgments (see Chapter 3.1.1). More recently, however, presuppositions have become the subject of quantitative work. There are experimental studies that span different presupposition triggers, languages, and ways of measurement (see e.g. Smith & Hall 2011, Xue & Onea 2011, Tonhauser et al. 2018, Tonhauser 2020, and the contributions in Schwarz 2015). This chapter contributes to this body of experimental work on presuppositions by investigating one particular type of presupposition, factivity, in Spanish. I present the results of a rating study on 44 Spanish clause-embedding verbs, which was closely modelled after Degen & Tonhauser's (2022) study on English.

The study makes use of (a Spanish translation of) the *certain that*-diagnostic for factivity, which is used in e.g. Tonhauser et al. (2018), Degen & Tonhauser (2021, 2022). This diagnostic assesses the degree to which a speaker is understood to be certain that the propositional content of the complement clause to some embedding verb is true, which is indicative of how likely the speaker is to presuppose that content. I go into detail on Degen & Tonhauser (2022) and this diagnostic in Section 4.1. The results of the present study on Spanish indicate that while some verbs show low or high average certainty ratings, there is a considerable number of verbs with intermediate mean ratings. When ordered by mean certainty rating, the verbs show a gradual cline in factivity, rather than abrupt differences between distinct groups of verbs. As result of this gradual cline, the differences between assumed verb types such as factive vs. non-factive only show as overall tendencies, without clear cut-off point between the types. For more fine-grained distinctions between further sub-types within these types, e.g. between semi-factives and true factives, it is even more difficult to pinpoint a distinction in the data, since these verbs show similar patterns overall. In this regard, my Spanish results parallel those of Degen & Tonhauser (2022) for English, who observe a similar gradualness in ratings, and who also find no evidence for abrupt cut-off points between assumed types or sub-types of embedding predicates.

While this study represents a standalone contribution to the study of factivity, it serves a more specific purpose in the greater context of this dissertation. The results make it possible to verify the assumptions that I made on the categories of the embedding verbs in my acceptability study on the complementizer *que* 'that' in globally embedded gapping (Chapter 3). Recall that in that study, the distinction between factive and non-factive verbs played a major role in accounting for the effect of *que* on acceptability. Further, that chapter featured the distinction of two sub-types within the factives, semi-factives and true factives, and two-sub-types within the non-factives. While the partition into the sub-types was driven by a grammatical factor independent to factivity (i.e. mood selection),[1] the initial separation into factive and

[1]Recall that I do not associate the distinction between semi-factive and true factive with any meaningful semantic or prag-

non-factive proceeded on the basis of my intuitive judgments. As I point out multiple times in Chapter 3, this procedure was not ideal due to my lack of practical knowledge of Spanish. As a consequence, my assessment of what verb belongs to which type may have been inaccurate for many of the embedding verbs. The study in this chapter illuminates whether and to what degree such false assessments occurred. To foreshadow the results, my categorization of the verbs in Chapter 3 was overall pretty accurate.

Furthermore, the results of the study open up a new way of looking at the experimental results of Chapter 3. In that study, I found that *que* in gapping led to a degradation in acceptability for some of the embedding verbs. The central hypothesis to account for this degradation was that it is determined by the factivity of the embedding verb: I hypothesized that non-factives show little or no degradation with *que*, while factives do show such a degradation. The acceptability patterns of Chapter 3 already provided some evidence against this hypothesis, since only verbs of saying and thinking (i.e. the indicative-taking sub-set of the non-factives) seemed to allow *que* with little or no degradation. Other non-factives patterned with the factives in showing a degradation. In the present chapter, I present a statistical comparison of the *que*-data from Chapter 3 and the certainty data, which provides further evidence against the factivity-based hypothesis. Certainty only correlates with the *que*-effect in gapping when the sample of verbs includes the verbs of saying and thinking. Without this verb type, the correlation disappears. This result illustrates the special role of verbs of saying and thinking (and not of non-factive verbs as such) in the *que*-effect in gapping, thus reinforcing the conclusions of Chapter 3.

This chapter is structured as follows. In Section 4.1 I describe Degen & Tonhauser's (2022) study, on which my study is based, in some more detail. Section 4.2 contains the details of the design, materials, procedure, and participant selection in my study. Section 4.3 presents and discusses the results. In Section 4.4, I relate the results of the study to the results of Chapter 3. Section 4.5 concludes.

4.1 How to test factive presupposition quantitatively (Degen & Tonhauser 2022)

Degen & Tonhauser's (2022) study contains the 20 English clause-embedding predicates in (1). Following the categories in the existing literature on factivity, Degen & Tonhauser (2022) assume that predicates fall into the categories non-factive and factive. Within the factives, the predicates separate into categorically factive and optionally factive predicates. This latter sub-type encompasses verbs that, according to introspective judgments in the literature, show a variable behaviour as regards factivity (i.e. they may receive both factive and non-factive interpretations). Categorically factive predicates (supposedly) do not show this variability; they always receive a factive interpretation.

(1) **Predicates tested in Degen & Tonhauser (2022: 559)**

a. **Non-factive**
pretend, say, suggest, think, be right, demonstrate

b. **Categorically factive**
be annoyed, discover, know, reveal, see

matic difference (see Chapter 3.1.1). As we will see in Section 4.3, this treatment of the factive sub-types has been most likely appropriate, since the results do not suggest a categorical difference between these two sub-types in terms of factivity.

c. **Optionally factive**
acknowledge, admit, announce, confess, confirm, establish, hear, inform, prove

This categorization is similar to the categorization I use in Chapter 3.1. There, I also assume the existence of the categories non-factive and factive, with two sub-types in each category. For the factives, the sub-types are semi-factives and true factives. These sub-types behave like Degen & Tonhauser's (2022) categorically and optionally factives, in that they show variable presupposition (semi-factives) and fixed presupposition (true factives). In Spanish, the partition of factives into semi- and true factives (per assumption) goes along with a split in mood selection: semi-factives select for the indicative, true factives for the subjunctive. The English mood system does not impose these strong selectional requirements on clause-embedding predicates, which is why mood is not relevant for English in this regard. This explains why the partition of non-factives into two types (indicative-taking vs. subjunctive-taking) is not necessary for English.[2]

Notice however that the individual predicates in Degen & Tonhauser's (2022) categorically and optionally factives do not align neatly with my true factives and semi-factives. The Spanish equivalents of some of their categorically factives (*saber, descubrir* and *ver*, the equivalents to *know, discover* and *see*, respectively) are in the semi-factives in my classification. However, the equivalents of their optionally factives that feature in my experiments are also classed as semi-factives (i.e. *admitir* 'admit' and *oír* 'hear'). A discussion of the potential reasons for this partially divergent classification is beyond the scope of this dissertation. Since the point of this chapter is not to compare different approaches to verb classification anyway, these differences may be safely ignored. As for Degen & Tonhauser's (2022) non-factives, two of them feature in my (indicative-taking) non-factives: *decir* 'say' and *pensar* 'think'. I provide a comparison of the verbs that are contained in both Degen & Tonhauser's (2022) study on English and my study on Spanish regarding their certainty ratings in Section 4.3.1.

To obtain a measure of how factive these verbs are, the experimental design must reflect two crucial properties of factive presuppositions: one, the inference that the speaker takes the propositional content of the complement clause to a given embedding predicate (the 'CC' for short) as given and uncontroversial in discourse; and two, the existence of this inference in entailment cancelling environments (see Chapter 3.1.1 for a discussion of both these properties). Degen & Tonhauser (2022) implement the first requirement by making use of the *certain that*-diagnostic (see also Tonhauser et al. 2018, Degen & Tonhauser 2021): participants are asked to indicate how certain they think someone who utters a sentence with clausal embedding is that the CC is true. High ratings for individual predicates indicate that the speaker is understood to be certain that the CC is true, i.e. that they presuppose the CC. If speakers are understood to presuppose the CC of some clause embedding predicate they use, one may conclude that this predicate is factive. However, this conclusion is only valid under the condition that the second relevant property of factivity holds as well.

This second property of factivity requires the certainty of the speaker that the CC is true to project from under entailment cancelling environment. Testing factivity thus requires that the embedded clauses

[2]Degen & Tonhauser (2022) actually do split their non-factives into two types: veridicals and non-veridicals. See Chapter 3.1.1 for a short explanation of veridicality. For the purposes of this chapter, veridicality is of no importance; I will thus not discuss this difference in non-factive predicates.

be presented within one such environment. In their experimental materials, Degen & Tonhauser (2022) chose polar questions as the entailment cancelling environment. An individual trial in their experiment thus involved the two units in (2): a polar question in a direct speech context like (2a), which contains the embedding predicate to be tested (here *discover*), and the response prompt in (2b) containing the *certain that*-diagnostic for the CC of the critical item. Given that the *certain that*-diagnostic asks participants to judge the certainty of the speaker (*Helen* in 2a), the item itself (2a) necessarily mentions the speaker explicitly as the subject of the direct speech.

(2) **Sample trial from Degen & Tonhauser (2022: 561)**

a. Helen asks: "Did Amanda discover that Danny ate the last cupcake?"

b. Is Helen certain that Danny ate the last cupcake?

Degen & Tonhauser (2022) tested these experimental materials in two experiments that implemented different ways of indicating a response to the *certain that*-diagnostic (i.e. two different ways for participants to respond to 2b). In one experiment (their Experiment 1a), participants indicated their response with a slider on a continuous scale that had the endpoints *yes* and *no*. This procedure allowed participants to give intermediate ratings. In the other experiment (their Experiment 1b), participants had to give categorical answers, i.e. they had to choose between *yes* and *no*, with no intermediate option.

Both these forms of measurement led to similar results in the two experiments, which I summarize here in a nutshell. While non-factives and factives overall receive the expected respective low and high certainty ratings, the two types do not form distinctly different categories. The difference between factives and non-factives was gradual rather than abrupt, which means that between the lowest rated predicates and the highest rated predicates, some predicates received intermediate ratings. Within the two factive sub-types (canonically and optionally factives), this gradualness was even more pronounced. This result provides problems for the categorization of factive predicates in terms of binary, mutually exclusive categories. As Degen & Tonhauser (2022: 564) put it: "a nonarbitrary line between canonically factive and optionally factive predicates cannot be drawn." My experiment on Spanish will provide comparable results (Section 4.3).

Degen & Tonhauser's (2022) study shows how a quantitative examination of factivity may be implemented by using the *certain that*-diagnostic. The results of the study have an impact on our understanding of factivity: the gradualness Degen & Tonhauser (2022) observe is unexpected from many treatments of factivity in the theoretical literature. I will not discuss the obvious question that arises from these results of how to account for the gradualness in the data. Instead, I will focus on questions that relate more directly to the topic of this dissertation: is the gradualness in factivity that Degen & Tonhauser (2022) observe in English cross-linguistically robust? If so, we expect similar results for an equivalent study in Spanish. And, more closely related to the general topic of this dissertation: has my categorization of the clause-embedding predicates of Chapter 3 been correct? And, can quantitative data on factivity in Spanish bring us closer to an explanation of the *que*-effect in gapping? I approach these questions in the rest of this chapter.

4.2 Experimental design, materials, procedure, and participants

My study consisted of one experiment, which tested the Spanish clause-embedding verbs in Table 4.1. There are 44 verbs in total, which means that this study is larger in scale than Degen & Tonhauser's (2022) study on English, which tested 20 embedding predicates. The verbs in the present study also featured in the experiments on the complementizer *que* 'that' in globally embedded gapping, see Chapter 3. That chapter (especially Section 3.1) details the categorization of these verbs into the four types you see in Table 4.1, i.e. the partition by mood selection (indicative vs. subjunctive) and factivity. Since the factivity of these verbs is the subject of the present experiment, their a-priori categorization into factive and non-factive groups is not appropriate. However, since one goal of this study is to see whether the assumptions I have made so far as regards these verbs have been accurate, I will maintain this categorization. In this context, the terms 'factive' and 'non-factive' function as descriptive labels rather than as actual claims about their semantic or pragmatic features. In a similar vein, I will continue to use the terms 'semi-factive' and 'true factive' to keep apart the indicative-taking and subjunctive-taking factives. Recall from the discussion of these terms in Chapter 3.1.1 that I do not associate any systematic semantic or pragmatic differences with these two factive sub-types. Their only difference lies in mood selection.

Table 4.1: Verbs of the experiment by their types (in the third person singular present indicative, like in the experimental items).

Non-Factives (with indicative)		**True Factives** (with subjunctive)		**Non-Factives** (with subjunctive)		**Semi-Factives** (with indicative)	
Spanish	*English*	*Spanish*	*English*	*Spanish*	*English*	*Spanish*	*English*
afirma	claims	admira	admires	desea	wishes	acepta	accepts
calcula	estimates	alaba	praises	desmiente	denies	admite	admits
cree	thinks	apoya	supports	duda	doubts	averigua	finds out
cuenta	tells	aprueba	approves	espera	expects	descubre	discovers
dice	says	critica	criticizes	niega	denies	ignora	ignores
imagina	imagines	defiende	defends	prohíbe	forbids	nota	notes
piensa	thinks	desaprueba	disapproves	propone	proposes	observa	observes
presume	presumes	detesta	detests	quiere	wants	oculta	hides
relata	relates	lamenta	regrets	rechaza	rejects	oye	hears
sospecha	suspects	odia	hates	recomienda	recommends	sabe	knows
supone	supposes	respeta	respects	teme	fears	ve	sees

With the help of a native Spanish speaker,[3] I constructed 22 clauses, which served as the complements to the embedding verbs. Two of these clauses are in (3) and (4).

(3) Lucía se gasta todo el dinero de su paga en eventos culturales.
Lucía REFL spend all her money of her pay on events cultural
'Lucía spends all her pocket money on cultural events.'

(4) Martina organiza su boda ella sola.
Martina organizes her wedding she only
'Martina organizes her wedding all by herself.'

[3]Thank you to Marta Cabezuelo for her help with the experimental materials. Also thank you to Jonilla Ried and Sven Weber for creating the German sentences that served as a starting-off point for the Spanish ones.

Each of the 22 clauses paired with each of the 44 embedding verbs to yield a total of 968 sentences. The heterogeneous selection requirements of the embedding verbs imposed a number of restrictions on the form and content of the embedded clauses. For the emotives (i.e. true factives), the content had to be such that the embedding subject could have a positive or negative attitude towards it (e.g. for the verbs *alaba* 'praises' vs. *odia* 'hates', respectively). The desideratives and the verb *prohíbe* 'prohibits' were only felicitous with complement clauses in the present tense, and one whose *Aktionsart* (Vendler 1957) was not a state (= an activity, accomplishment or achievement). Furthermore, the verb *calcula* 'estimates' called for a quantity expression in the embedded clause, like *todo el dinero* 'all the money' in (3), or an expression that could plausibly receive a quantity construal such as *su boda* 'her wedding (= the entirety of the event)' in (4).

As a result of the restrictions that some individual verbs imposed on their complement clauses, it was not feasible to include all 48 verbs from the gapping experiments of Chapter 3 in this experiment. For instance, the verb *considera* 'considers' requires an evaluative component or expression (e.g. an adjective) in its complement clause. To illustrate this requirement, let me make a brief comparison with the German translation equivalent of *considera*, *finden* 'consider' (literally 'find'). This verb shares the requirement of *considera* for an evaluative expression, as for example in *Ich finde, dass Maria *(gut) singt*, literally 'I find that Maria sings well'. Constructing the complement clauses in a way that satisfies this requirement of *considera* would have led to clashes with the requirements of some of the other embedding verbs. Consequently, I chose to not include *considera* in the sample. To keep the four verb types balanced, one verb of each of the other three types similarly did not enter the experimental materials. The non-included verbs are the aforementioned *considera* 'considers' (indicative-taking non-factive), *demanda* 'demands' (subjunctive-taking non-factive), *conoce* 'knows' (semi-factive), and *consiente* 'tolerates' (true factive). The choice as to which verb should be excluded proceeded on the basis of my native speaker informant's judgments as to which verb were causing the most problems in the materials, i.e. which verb sounded unnatural in the largest number of test sentences.[4]

The matrix and embedded subjects of the sentences were common Spanish first names. All embedded and matrix clauses had unique subjects. For example, the sentence in (4) above invariably came with the embedded subject *Martina* and the matrix subject *Marco*, independently of the embedding verb, see (5). Half of the embedded and matrix subjects were female names, and the other half were male names.

(5) Marco {afirma / calcula / admite / sabe / … } que Martina organiza su boda ella sola.
Marco claims estimates admits knows that Martina organizes her wedding she only
'Marco claims/estimates/admits/knows that Martina organizes her wedding all by herself.'

To facilitate the investigation of factivity, it was crucial to present these sentences within an entailment cancelling environment. To this end, I followed Degen & Tonhauser (2022) in presenting the items as polar questions. Since factive presuppositions are tied to the epistemic state of the speaker, each item was preceded by another personal name, which represented that speaker in the conversation context. This name was invariably *María* in all items. A sample of what these items looked like with the speaker and the polar question is in (6). In this item, the crucial embedding verb is *afirma* 'claims'.

[4]Due to the large number of embedding verbs, it is likely that not all resulting sentences were perfectly natural. My native speaker informant made sure that no sentence was outright unacceptable.

(6) María: ¿Marco afirma que Martina organiza su boda ella sola?
María Marco claims that Martina organizes her wedding she only
'María: Does Marco claim that Martina organizes her wedding all by herself?'

In addition to these critical items, the materials featured six control items. These control items were polar questions without clausal embedding, see (7) for one of them. The ratings for the control items served as a baseline for uncertainty/non-projection. I return to this issue below.

(7) María: ¿Eugenio alimenta a sus gatos con carne ecológica de una tienda especializada?
María Eugenio feeds DOM his cats with meat organic from a shop specialized
'María: Does Eugenio feed his cats with organic meat from a specialty shop?'

Participants were instructed to imagine a situation in which they were at a party. The host of the party was their friend María, who is a very chatty person and loves talking about other people. Upon entering the kitchen, the participant overheard María uttering the question in the respective test sentence. In the experiment, participants then indicated how certain they thought María was that the CC was true, see (8) for the exact question for the item in (6). This question constitutes a Spanish version of the *certain that*-diagnostic.

(8) ¿Está María segura de que Martina organiza su boda ella sola?
Is María certain of that Martina organizes-IND her wedding she only
'Is María certain that Martina organizes her wedding all by herself?'

Participants indicated how certain they thought María was regarding the CC by dragging a slide ruler on a continuous horizontal scale. The extremes of the scale were labelled *sí* 'yes' and *no* 'no'. This procedure parallels that of Degen & Tonhauser's (2022) Experiment 1a (see Section 4.1). The output coding ranged in integers from 1 (slider position all the way at *no*) to 100 (slider position all the way at *sí*). For the control items, participants indicated how certain they thought María was regarding the content of the whole question, see (9).

(9) ¿Está María segura de que Eugenio alimenta a sus gatos con carne ecológica de una
Is María certain of that Eugenio feeds DOM his cats with meat organic from a
tienda especializada?
shop specialized
'Is María certain that Eugenio feed his cats with organic meat from a specialty shop?'

The complement clause of the diagnostic question invariably came in the indicative (e.g. *organiza* 'organizes' in 8 above, rather than subjunctive *organice*). In choosing this mood over the subjunctive, I followed the advice of my native speaker informant on what seemed the most natural choice for the experimental materials and the hypothetical situation the sentences were presented in. However, the subjunctive would have been a possible choice as well. Certainty predicates like *segura* show an alternation in mood selection similar to what I described in Chapter 3.1.2. It is therefore possible, that the choice of the indicative in the diagnostic influenced the certainty ratings. The results of the experiment do not provide any reason to suspect that the choice of the indicative in the diagnostic was entirely inappropriate. However, I will come back to this issue in the discussion of the certainty ratings for the main clause controls in Section 4.3 (especially Footnote 7).

I split the 968 test sentences into two blocks. Each block contained all the sentences for half of the embedding verbs. I then distributed the sentences in each block into 22 lists, so that each list contained each embedding verb (in the respective block) and each embedded clause exactly once (= Latin square). With the six control items, each participant thus judged 28 items per list.

Both blocks were web-based questionnaires hosted on SoSciSurvey.de (Leiner 2024). In the questionnaires, experimental items and fillers were presented one by one on separate pages. These pages contained the test sentence followed by the scale on which participants gave their judgments. I used Prolific.com to recruit participants. On this platform, the questionnaire was available to native Spanish speakers from Spain, who indicated that they grew up speaking their native language only. The first 300 participants of Block 1 had participated in at least one of the experiments from Chapter 3. I allowed this specific set of participants to participate in the experiment because they had performed well in their previous experiments (i.e. they were not excluded by the exclusion criteria in those experiments, see Chapter 3.2). In addition to these 300 participants, I collected the data of 30 more participants, who had not previously participated in any of my experiments.

To ensure the quality of the data, I set three exclusion criteria for the responses of the individual participants. These criteria are listed in (10). The first criterion excludes participants who gave high certainty ratings for the content of the non-embedded controls. This criterion relates to the assumption that the content of a polar question without clausal embedding should not be projective at all. This means that if a speaker (i.e. *María* in the experiment) utters a question whether *p*, the answer to whether María is certain that *p* should be very close to *no*. The second criterion excluded participants who gave low ratings across the board (i.e. for experimental items with embedding and main clause controls). This criterion relates to the assumption that for the CC of at least some of the embedding verbs, certainty ratings should be high (i.e. for factives). Since participants judged both (assumed) non-factives and factives in the experiment, I set the threshold for Criterion 2 so low as to only exclude participants who gave ratings that were entirely incongruent with this assumption, i.e. participants who basically only gave low certainty ratings. Criterion 3 is a combination of the first two criteria. It excluded participants who did not differentiate their certainty ratings, i.e. those who only gave 'middle-of-the-road' judgments. The reasoning for this criterion was that there is at least some variance to be expected in the ratings, which follows from the assumptions relating to Criteria 1 and 2.

(10) **Exclusion criteria for participants**

1. The mean rating for main clause controls was too high. (Threshold: max. 50)
2. The mean rating for experimental items was too low. (Threshold: min. 20)
3. The participant used too little of the available scale. (Threshold: min. difference of 50 between highest and lowest rating, i.e. use of at least half the available scale.)

In Block 1, Criterion 1 lead to the exclusion of 48 participants, criterion 2 excluded 5 participants, and criterion 3 excluded 11 participants (8 of which were also excluded by one of the other two criteria). Thus, from the total 330 participants that had completed the questionnaire, 274 participants remained after exclusion. After two months, those participants that had not been excluded in Block 1 were asked to participate in Block 2. In total, 256 participants completed the questionnaire of Block 2. Of these, 18

participants were excluded by Criteria 1 to 3 (15, 1, and 2 participants, respectively). In total, the data of 238 participants entered the analysis. This means that I have that number (238) of observations per embedding verb, and twice that number (476) per embedded clause. For each individual test sentence (i.e. matrix clause-embedded clause combination like 6 above) there are between 9 and 14 judgments (median: 11). In total, there are 10472 observations for experimental items. For the main clause controls, the total number of observations is 2856 (476 per control item, i.e. one per participant in each block).

The relatively high number of participants excluded by Criterion 1 (14.5% of all participants in Block 1), i.e. due to too high mean ratings for the controls, is somewhat concerning. A comparable criterion in Degen & Tonhauser's (2022) study excluded considerably fewer participants (5.3%).[5] I see two possible reasons for the high rejection rate in my experiment: first, the indicative mood in the diagnostic (i.e. in the complement clause to *segura*) might suggest a relatively high certainty, which is a possibility I noted above.[6] I return to this possibility in the discussion of the certainty ratings for the main clause controls in Section 4.3 (especially Footnote 7). Second, the translation of the English phrase *certain that* to Spanish *segura de que* may not have resulted in a perfectly equivalent diagnostic. Depending on the exact meaning of the adjective *segura*, participants might be more likely to indicate higher ratings for sentences in Spanish than would be recorded for an equivalent English sentence with the *certain that*-diagnostic, even if the actual level of certainty is identical in the two sentences. Of course, I chose to translate *certain* with *segura* in the belief that these two terms are equivalent in this regard, and I have no reason to suspect that these two terms are not equivalent. All in all, I have no evidence for or against this possible explanation for the high rejection rate from Criterion 1, and I generally see no way of completely resolving this problem. All I can do here is ignore this issue and proceed.

4.3 Results by verbs and verb types

This section contains a summary and discussion of the certainty ratings. I start by discussing the collapsed data for the four verb types I assumed in Chapter 3, and then turn to the individual embedding verbs. Table 4.2 and Figure 4.1 contain the mean certainty ratings for the four verb types and for the main clause controls. As expected, the mean rating for the main clause controls is relatively low.[7] Also as expected, the two non-factive types received considerably lower mean ratings than the two factive types. This result means that on the whole, my categorization of the verbs into non-factive and factive types was accurate. However, these two types received noticeably higher certainty ratings than the main clause

[5] See Supplement C to their article (Degen & Tonhauser 2022) for details on their exclusion criteria. The supplementary material is available at https://muse.jhu.edu/article/865005 (accessed 24 July 2024).

[6] Note, however, that it is unclear how exactly this influence of mood selection would work in this context. A grammatical property like mood of some question Q about some person X's certainty towards the truth of a proposition p (i.e. Q = *Is X certain that p?*) should, intuitively speaking, at most influence the interpretation of the utterer of Q's certainty of p and not the interpretation of X's certainty of p.

[7] A considerable number of participants was excluded from analysis due to their main clause ratings being too high (mean >50, see Criterion 1 in the list in 10 above). This naturally raises the question of whether the low mean rating for the controls is an artifact of this exclusion procedure, which lead to participants with high main clause ratings systematically missing from the data. However, this does not seem to be the case: main clause controls show similarly low ratings when the data from all participants enter analysis, including those that indicated high main clause ratings. In this case, the mean rating for the controls is still lower than that of any of the four verb types, and also lower than that of any of the individual verbs (cp. Figure 4.2 below).

controls, which is unexpected under the assumption that non-factives should not project their CC at all and thus pattern like the main clause controls.

Table 4.2: Mean certainty ratings per verb type and main clause control items (*N* = number of observations, *SD* = standard deviation, *SE* = standard error, *CI* = 95% confidence interval)

Factivity	Mood	N	Mean rating	SD	SE	CI
Non-factive	Indicative	2618	39.36	32.99	0.64	1.26
Non-factive	Subjunctive	2618	46.05	35.85	0.70	1.37
Semi-factive	Indicative	2618	77.31	28.21	0.55	1.08
True factive	Subjunctive	2618	82.51	24.44	0.48	0.94
Main clause controls		2856	18.25	25.81	0.48	0.95

Furthermore, there is a small but noticeable difference between the indicative-taking verbs and the subjunctive-taking verbs. Within both non-factives and factives, the subjunctive-taking the subgroup receives slightly higher certainty ratings than the indicative-taking subgroup. To gain a deeper insight on this (apparent) effect of mood selection, I used `R` (R Core Team 2023) and the `afex`-package (Singmann et al. 2022) to fit a linear mixed effects model on the certainty ratings. The model contained the two two-level fixed factors factivity (non-factive vs. factive) and mood selection (indicative vs. subjunctive). (The data for the main clause controls did not enter the model). To facilitate the interpretation of interaction effects, the fixed factors had a sum-to-zero contrast coding. The random effects structure contained random slopes and intercepts for both factors for participants and items (= the 22 embedded clauses). As expected, the model indicated a statistically significant effect of factivity ($b = 18.61, SE = 0.50, t = 36.95, p < 0.001$). Crucially, the effect of mood was significant as well ($b = 2.97, SE = 0.41, t = 7.32, p < 0.001$). The interaction was not significant ($b = -0.37, SE = 0.26, t = -1.42, p = 0.16$).

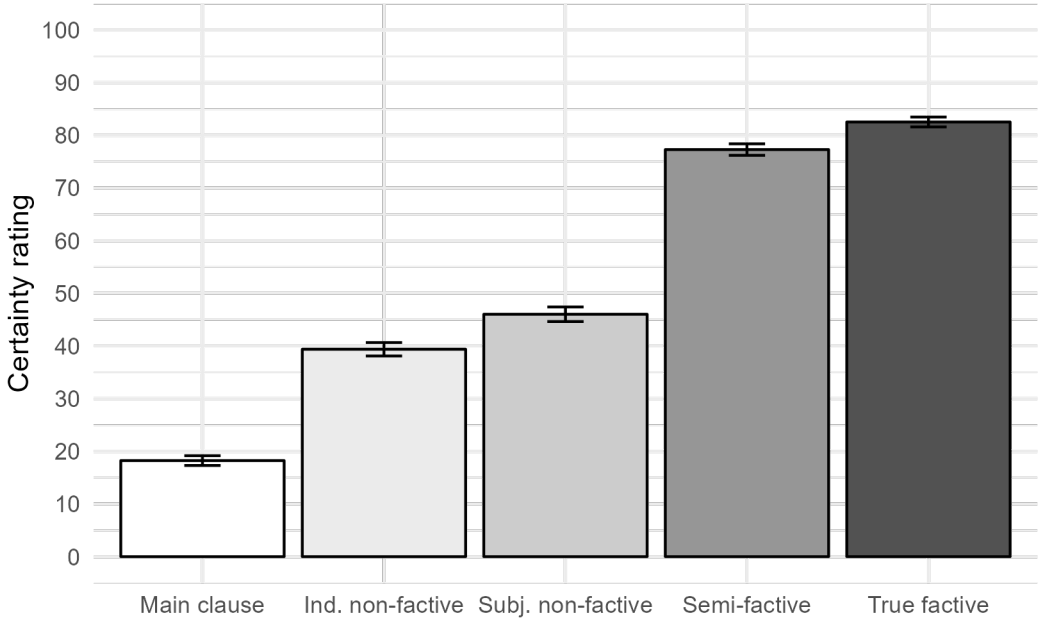


Figure 4.1: Mean certainty ratings and 95% confidence intervals for the four verb types and main clause controls.

The standard conclusion from the statistically significant mood effect would be the inference that the pattern in these data represents a genuine underlying pattern for Spanish mood generally. This conclusion entails that the subjunctive is categorically more projective than the indicative. The absence of a significant interaction suggests that this difference obtains equally within the (presumed) non-factive and factive sub-types. Recall that in the discussion of these four verb types in Chapter 3.1.1, I predict no such difference between factive and non-factive sub-types; the only difference I predict is the difference between non-factive and factive as such. For the (presumed) factives specifically, there thus seems to be a categorical semantic difference between the semi-factives and the true factives in terms of presupposition projection – an option I explicitly reject in Chapter 3.1.1.

While I cannot argue with the fact that the difference in mood is statistically significant in my data, the size of the effect itself casts doubt on the conclusion that the subjunctive is categorically associated with higher presupposition projection. The difference between the indicative- and subjunctive-taking subgroups (= the mood effect) is very small: 6.7 (out of a possible 100) for the non-factives, and 5.2 for the factives. With this small effect of less than 1/10 of the available scale, I would like to argue that it would be misleading to claim a categorical difference in factivity between the two moods. While I acknowledge the significant *p*, I reject this strong version of the inference based on the small effect itself. I suspect that the significant mood effect is a result of my selection of individual verbs (in combination, perhaps, with the relatively large sample size).

This leaves the apparently strong effect of factivity on the certainty ratings. The difference between the verbs I classed as non-factive and factive is much bigger than between the moods, and thus seems to represent a genuine categorical difference between these verb types. This conclusion is good because it validates my assumptions as to the factive or non-factive category membership of the respective verbs. However, a closer look at the ratings for the individual verbs casts doubt on a fundamental prerequisite of this conclusion, i.e. the assumption of categorically distinct factive and non-factive categories. Figure 4.2 contains the mean certainty rating for each of the 44 embedding verbs and the main clause controls. Those verbs whose English equivalents also feature in Degen & Tonhauser's (2022) study are marked with a star. I will provide a short comparison of English and Spanish in Section 4.3.1; for now I focus on the results of my experiment.

Like for the collapsed data with the four verb types, the main clause controls received the lowest mean rating. All verbs received mean certainty ratings that were higher than those of the main clause controls. This result seems to suggest that the fact that a structure involves clausal embedding alone creates the possibility for projection of the CC, independently of the exact embedding verb.[8] Nevertheless, there are large differences in certainty ratings between the verbs. The verb with the lowest average certainty is *propone* 'proposes', with a mean rating of around 26. The verb with the highest certainty rating is *sabe* 'knows' at around 92. Generally speaking, the two (presumed) non-factive types represented by the two lighter shades tend towards the left with lower values, while the (presumed) factive types represented in the two darker shades tend towards the right with higher values. Descriptively speaking, two verbs mark the "border" between non-factive and factive: the subjunctive-taking non-factive *duda* 'doubts' and

[8] Degen & Tonhauser (2022) find a similar result for English, which suggests that the possibility of projection by embedding alone, independently of the exact embedding predicate, may be a cross-linguistic phenomenon. More research is needed to determine the exact extent of this phenomenon, and to shed light on why this might be.

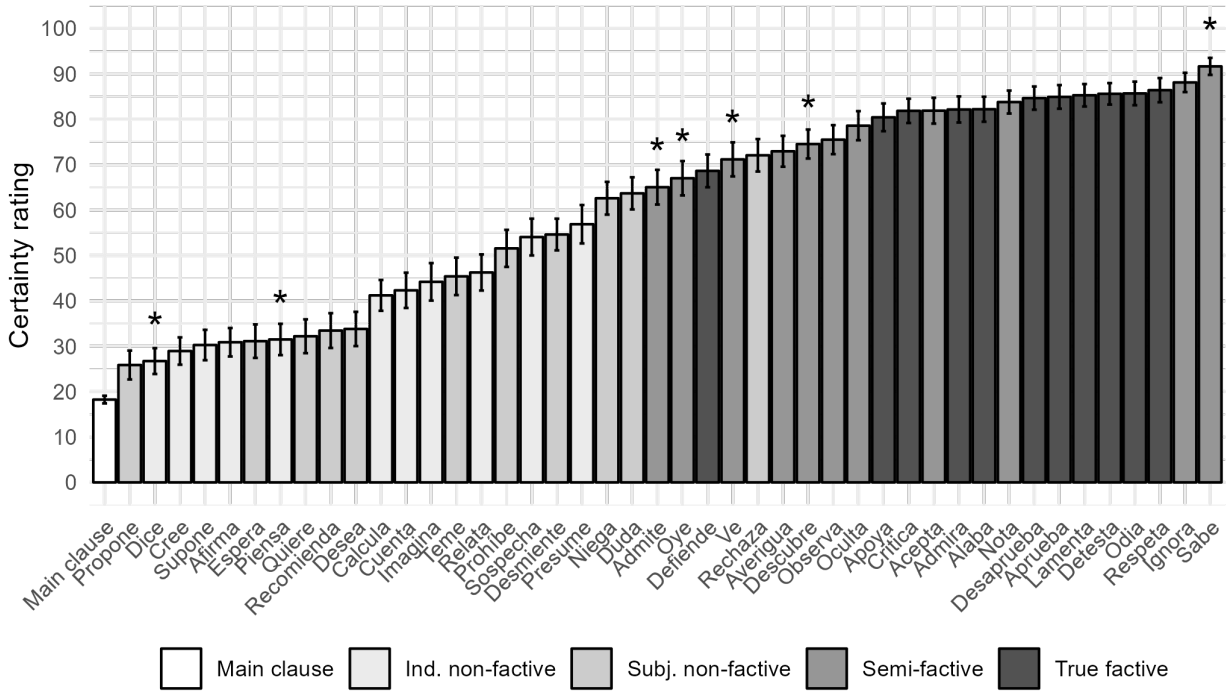


Figure 4.2: Mean certainty ratings and 95% confidence intervals for the 44 verbs and main clause controls. Stars indicate verbs whose English equivalents feature in Degen & Tonhauser's (2022) study.

the semi-factive *admite* 'admits'. The only apparent exception to this partition is *rechaza* 'rejects', which I classed as non-factive, but which is located in the right half, surrounded by (presumed) factives. This verb thus represents the most striking candidate for a misclassification on my part.

Despite these general observations, the most striking pattern in Figure 4.2 is the gradualness between the verbs. The differences between the assumed non-factives and factives only surface as tendencies; there are no clear-cut boundaries between these types. The actual difference between the means of the two aforementioned "border" verbs *duda* and *admite* is at a very small 1.4. Thus, even if we allow for the option that I categorized some of the verbs incorrectly (as I hypothesized in Chapter 3), the gradualness in certainty ratings I observe here casts general doubts on the assumption that there are two distinct classes of factive vs. non-factive verbs.[9]

The same line of reasoning extends to the distinction between the two factive sub-groups. While the statistical results above indicated a slight but significant difference between semi-factives and true factives on the whole, for the individual verbs no such split between the factive sub-types is apparent. Many of the true factives receive comparatively high ratings within the factives, but the two highest rated verbs are semi-factives: *sabe* 'knows' and *ignora* 'ignores'. Conversely, one of the true factives, *defiende* 'defends' ranges towards the lower end within the factive verbs. The possibility once more arises that I miscategorized these verbs within the factive sub-groups. This means that *defiende* may actually be semi-factive, while *sabe* and *ignore* may actually be true factive. (For *sabe* 'knows', this is unlikely since in the

[9]Despite the conclusion that the assumption of distinct verb types as concerns factivity is likely inaccurate, I will maintain the conclusion I drew above that my categorization of the verbs was largely correct, at least within its underlying, and as it has now turned out partially inaccurate, assumptions.

literature on English, *know* and knowledge verbs generally are treated as prime examples of semi-factivity, see Chapter 3.1). However, even if we assume such a miscategorization, there is still the gradualness in the data that seems incompatible with distinct verb types. As for the mean ratings for all 44 verbs, we observe a gradual cline in certainty ratings within the factives. Independently of my own categorization, we do not observe any abrupt difference between two sub-types that one would expect if semi-factives and true factives were indeed two entirely distinct categories anywhere within the higher rated verbs.

A similar gradualness is at play for the (presumed) non-factives. Again, the subjunctive-taking non-factives do not form a joint class to the exclusion of the indicative-taking non-factives, and the differences between verbs are gradual rather than categorical. Between the verbs *desea* 'wishes' and *calcula* 'estimates' there is a noticeable gap (with non-overlapping confidence intervals), which suggests that there might be a categorical difference between verbs to the left and right of this gap. However, this option seems unlikely given that this gap is in the middle of my presumed non-factives, with verbs of both assumed sub-types to either side. I see no reason to assume that this gap represents a genuine line of demarcation between types of embedding verbs. More research is needed to investigate this apparent gap, but for now I will assume that there exist clause-embedding predicates with mean certainty ratings within the range of that gap, which my experimental materials do not feature by coincidence. I leave the question of which embedding predicates may fill this gap to future research.

One pattern in the data that might indicate a genuine difference between verb types concerns the subjunctive-taking non-factives. Recall that this is a heterogeneous category, containing desideratives on the one hand, and verbs with implicit negation on the other (see Chapter 3.1). Within this category, the negation verbs received higher ratings across the board than the desideratives, as shown in Figure 4.3. The desideratives all received comparably low ratings, and range with the lowest rated verbs overall in the experiment. All negation verbs received higher ratings than the desideratives, and collectively show a more pronounced incline in certainty ratings. The difference between the highest rated verb *rechaza* 'rejects' and the lowest rating verb *teme* 'fears/worries' is at about 27.[10] The ratings for the negation verbs suggest that the semantic class of negation verbs might genuinely be more projective than other verbs that I classed as non-factive. In other words, the categorization of this class of verbs as non-factive may have been systematically incorrect. I leave an in-depth investigation of the factivity of negation verbs to future research.

This potential misclassification of negation verbs as a semantic class (as opposed to the potentially incorrect categorization of individual verbs) may shed new light on the results of Chapter 3. Recall that the subjunctive-taking non-factives show a degradation of *que* in globally embedded gapping, while the indicative-taking non-factives show no such degradation. If the subjunctive-taking non-factives now contain a group of verbs that are actually likely to receive a factive interpretation (i.e. the negation verbs), the high degradation of subjunctive-taking non-factives with *que* may thus still be tied to factivity. This possibility goes against the comparison of the indicative-taking and subjunctive-taking non-factives above,

[10] Footnote 4 in Chapter 3.1 explains why *teme* 'fears/worries' is classed as a negation verb. I repeat the content of that footnote here for convenience: The reason for classing *temer* as a negation verb is not immediately obvious from its English equivalents 'fear' and 'worry'. However, the Latin origin of this verb *timeō*, indeed seemed to carry such a negative component to its meaning and selected for the negative complementizer *ne* instead of unmarked *ut* (Gildersleeve & Lodge 1905: 237, §550). For this reason I will group *temer* with the negation verbs within the subjunctive-taking non-factives. (Unlike *temer*, the Latin distinction between negative and non-negative complementizers did not survive in Spanish.)

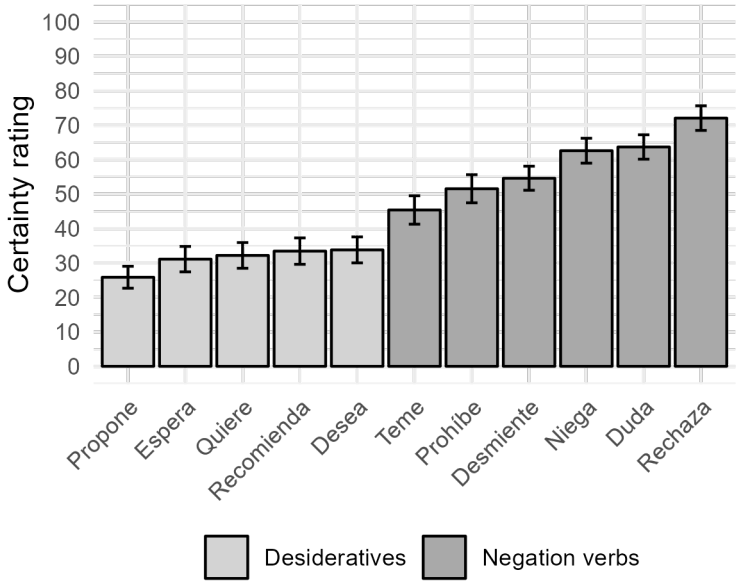


Figure 4.3: Zoom into the two sub-types of the subjunctive-taking non-factives (mean certainty ratings with 95% confidence intervals).

where these two types do not show the large difference between certainty ratings one would expect from the results of Chapter 3. To clear up this confusing and somewhat incoherent situation, I test the correlation of the certainty ratings of the individual verbs with their *que*-degradation in globally embedded gapping obtained in Chapter 3 in Section 4.4.

Before I get to that discussion, I take a short detour via two sets of individual verbs that do not so neatly align with any of the types or sub-types that I assume. The first set of verbs are those that Rivero (1971) discusses concerning the interplay of mood selection and factivity (see Chapter 3.1.1). The second set of verbs are those that feature in both my and (in their English equivalents) in Degen & Tonhauser's (2022) study (see Section 4.1). The discussion of these two sets of verbs is not crucial to the rest of this chapter.

4.3.1 Relating the results to claims and observations in the literature (Rivero 1971, Degen & Tonhauser 2022)

In Chapter 3.1.2, I discuss variability in mood selection within and across embedding predicates, concentrating on the types and sub-types of verbs that I work with in that chapter. Towards the end of that discussion I also briefly touch upon a group of verbs discussed by Rivero (1971: 307 f.), which show alternating mood selection and which are not contained within a single specific one of my verb (sub-)types. These verbs are *admitir* 'admit' (a semi-factive in my classification), *esperar* 'hope/expect' and *negar* 'deny' (subjunctive-taking non-factives), and *sospechar* 'suspect' and *suponer* 'suppose' (indicative-taking non-factives). For these verbs, Rivero (1971) claims that the indicative triggers a factive reading and the subjunctive a non-factive reading. She explicitly claims this to be the case in the entailment cancelling environment of polar questions (amongst other such environments). Since I tested polar questions in my experiment, the data on these five verbs may be adduced to check Rivero's (1971) claim as to the interplay of mood choice and factivity.

In my experimental materials, these verbs only came in combination with one mood, i.e. either the indicative or the subjunctive. It will therefore not be possible to discern the impact of mood choice on factivity for the individual verbs from the data in my experiment. However, the prediction that follows from Rivero's (1971) claim is that all of these verbs show low ratings with the subjunctive and high ratings with the indicative. To facilitate the examination of this prediction, Figure 4.4 contains the certainty ratings for Rivero's (1971) five verbs. The height of the bars (i.e. the mean certainty ratings) are identical to those in Figure 4.2 for the respective verbs, but the colour of the bars now codes mood selection and not verb type.

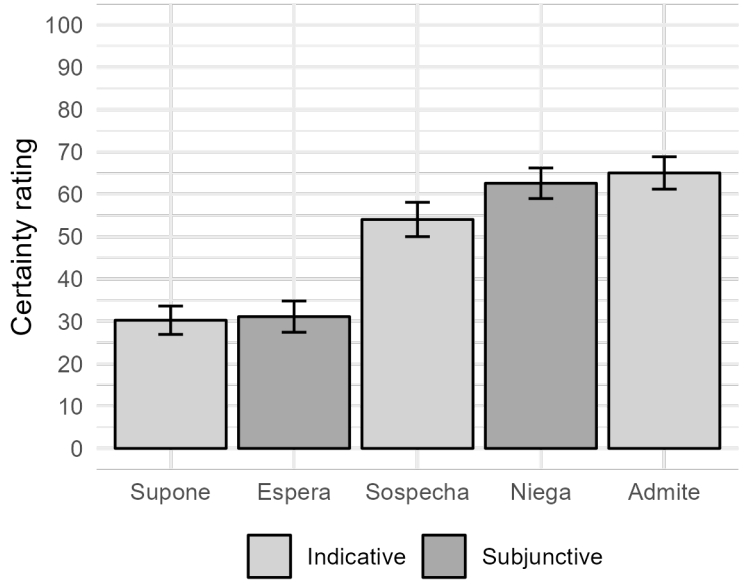


Figure 4.4: Verbs with variable mood selection according to Rivero (1971: 307 f.) (mean certainty ratings with 95% confidence intervals), and the respective mood they appeared with in the experiment.

Upon visual inspection there seems to be no correlation between factivity and mood selection: with the indicative, ratings may be either relatively high (as for *sospecha* and *admite*) or relatively low (as for *supone*). The same pattern obtains for the subjunctive: with this mood, we also see a relatively high rating for one verb (*niega*) and a relatively low rating for another (*espera*). Thus, while there may well be a difference in interpretation for these verbs depending on whether they select the indicative or the subjunctive (as claimed by Rivero 1971), the difference is unlikely to lie in factivity. The exact nature of the contribution of variable mood selection to meaning is an issue for future research. All in all, these patterns in Rivero's (1971) verbs strengthen the conclusion I reach above on the basis of the data from the verb types: mood does not seem to determine factivity, even in those verbs that show variable selection. This conclusion may be problematic for approaches that assume a tight connection between mood and factivity, which does not only apply to Rivero (1971), but also to more recent proposals (e.g. Quer 2011, Meinunger 2017, Giannakidou & Mari 2019).

I now turn to the comparison of my Spanish data and Degen & Tonhauser's (2022) data on English.[11]

[11]Specifically, I compare my results to those of Degen & Tonhauser's (2022) Experiment 1a, which like my experiment implemented a continuous scale for the response variable. The data from Degen & Tonhauser's (2022) Experiment 1b, where participants gave categorical responses, is not relevant here. For details on Degen & Tonhauser (2022), see Section 4.1.

There are seven pairs of translation equivalent verbs in the two studies, see Table 4.3.[12] One question that these verb pairs may shed light on is whether the translation equivalence across the two languages only obtains for the lexical semantic components of the verbs, or whether it also holds in further meaning components such as the interpretation of speaker certainty in polar questions. The data from the verb pairs suggests that the latter is the case: those verbs that received low certainty ratings in English also received low ratings in Spanish, and those verbs that received high ratings in English also received high ratings in Spanish.

Table 4.3: Mean certainty ratings and residuals in the linear model (see Figure 4.5) for translation-equivalent verbs in Degen & Tonhauser's (2022) study on English and my study on Spanish.

English		Spanish		Residuals
Verb	Rating	Verb	Rating	in Figure 4.5
say	24.46	dice	26.73	2.25
think	19.52	piensa	31.49	-8.11
admit	65.59	admite	65.03	-0.26
hear	74.70	oye	67.01	6.60
see	80.93	ve	71.16	8.10
discover	77.85	descubre	74.54	1.17
know	86.42	sabe	91.65	-9.75

The correlation between the English and Spanish certainty ratings for the individual verbs is visualised in Figure 4.5. It shows the mean certainty rating for Spanish on the horizontal axis and the mean certainty rating for English on the vertical axis. I used `R` to fit a linear model to the data from these seven verb pairs.[13] The model indicated a positive, statistically significant correlation of the English and Spanish certainty ratings ($b = 1.14$, see the dashed regression line in Figure 4.5, $p < 0.001$). The slope of the model being close to 1 indicates that not only do the values correlate, but they correlate nearly one to one. This finding suggests that the verbs in the sample are on the whole not only equivalent in their lexical semantics, but also in their projection behaviour. It remains to be seen in future research whether and in what way these two aspects of meaning (lexical semantics and projection) are dependent of each other.

Despite the strong and statistically significant correlation, some of the individual verbs show minor deviations from the regression line in Figure 4.5. Table 4.3 above contains a column with the value of the residual for each verb pair (i.e. the measure of distance along the vertical axis between the individual points in Figure 4.5 and the regression line). The verb pairs with positive residual values received higher average certainty ratings in English, and the verb pairs with a negative residual values received higher average certainty ratings in Spanish. In this latter direction, two verb pairs stand out: *know/sabe* and

[12] In Table 4.3, I transformed Degen & Tonhauser's (2022) data from a 0-1 scale to a 0-100 scale to fit with the results of my study, and calculated the mean certainty ratings with the function `SummarySEwithin()` from the package `Rmisc` (Hope 2022) for `R`.

[13] To be explicit about the statistical model: this was not a *mixed* model, i.e. it did not contain any random effects. The absence of random effects stems from the fact that the data do not contain any repeated measures (i.e. multiple measurements from the same participant or item). The mean values of certainty and *que*-degradation were calculated on the basis of data with repeat measures, but are themselves treated as individual measures. I dwell on this point since these statistical properties differentiate the model in question from the linear models in all other chapters and appendices of this dissertation, which are mixed models and thus do contain random effects. The present chapter contains two more non-mixed models in Section 4.4.

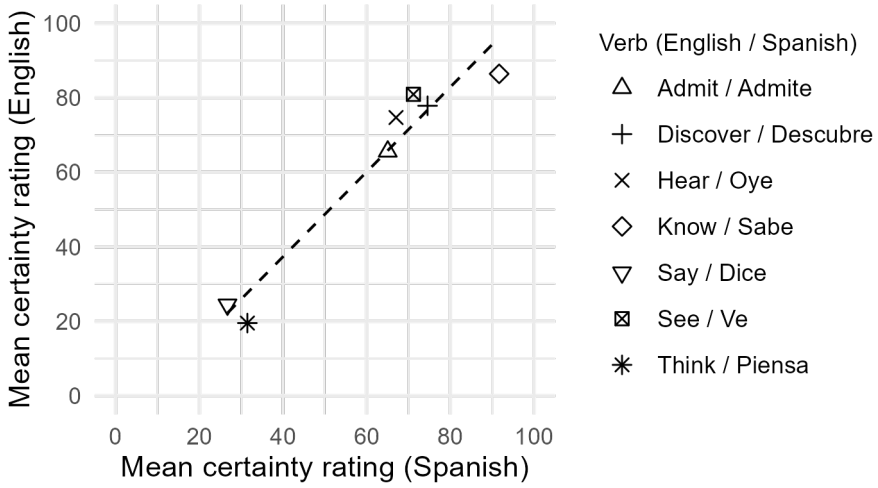


Figure 4.5: Correlation of mean certainty ratings of translation equivalent verbs from Degen & Tonhauser's (2022) study on English and my study on Spanish. The dashed regression line stems from a linear model.

think/piensa. These two pairs have the highest absolute residual values at around 10 and 8, respectively. The highest positive value is also at about 8, for the pair *see/ve*. Notice however that the respective distances of these verb pairs from the regression line in Figure 4.5 appear to be small. More research is necessary to see whether the deviations from the regression line for these verb pairs are indicative of a genuine difference in meaning or usage between English and Spanish, or whether these differences are coincidental.

Notice furthermore that the verb pairs in the sample happen to have mean certainty ratings that are either low (i.e. the two pairs *see/ve* and *think/piensa*), or cover a range along the higher end of the scale (all other pairs). In other words, there is a distinct lack of verbs in the sample with mean certainty ratings in the intermediate range. No verb has a mean certainty rating between 32 and 65 in either language. Verbs with such intermediate ratings are of special interest since intermediate certainty, and, by extension, intermediate factivity, are unexpected from the majority of theoretical accounts. Future comparative work should thus include verbs with such intermediate ratings.

4.4 Correlating certainty ratings and *que*-degradation in gapping

In this section, I connect the factivity data from the experiment in this chapter to the acceptability data from Chapter 3. On the basis of the individual verbs, I investigate whether the certainty ratings correlate with the degradation of *que* in globally embedded gapping. The motivation for this procedure stems from my analysis of the gapping data in Chapter 3.5. As a part of the analysis, I hypothesize that *que* in gapping is indicative of a non-factive structure, which leads to degraded acceptability ratings for verbs that usually are interpreted as factive. Given that both the certainty ratings from this chapter, and the effect of *que* on acceptability are gradual across the different verbs, I may reformulate this hypothesis in gradual terms: the more factive a verb is, the higher its degradation with *que* is expected to be. This hypothesis thus predicts a positive correlation of mean certainty ratings and the average *que*-degradation.

The mean certainty ratings and the average *que*-degradation for the 44 verbs are plotted in Figure 4.6. The vertical axis shows the mean certainty ratings. The vertical spread of the data points are thus identical to the height of the individual bars in Figure 4.2 above. This means the further along the vertical axis a verb is located, the higher the mean certainty rating in the experiment of this chapter. The horizontal axis shows the average degradation with *que* in globally embedded gapping (in comparison with the version without *que*). These values correspond to the bar heights in Figure 3.9 in Chapter 3.4. The further to the right a verb is located, the higher its average degradation with *que*. As you can see, there is one verb, *imagina* 'imagines', with a slightly negative value on the horizontal axis, which means that this verb on average was rated slightly more acceptable with *que* than without. To test whether the two measures are correlated, I fitted a linear model that predicted the average certainty rating per verb from its average degradation with *que* in gapping. The right side of Figure 4.6 shows the regression line that resulted from this model. The line has a positive slope ($b = 1.52$), which indicates that as predicted, verbs with a higher certainty rating show a higher degradation of *que*. The model indicated that this correlation is statistically significant ($p = 0.02$). However, the model indicated a low correlation coefficient (adjusted $R^2 = 0.11$), which means that despite the significant correlation of average certainty with *que*-degradation, certainty can account for only a relatively small part (11%) of the variation within the *que*-degradation data. This lack of explanatory power also shows in the relatively large residuals in Figure 4.6, which range from -37.2 to 35.3.[14]

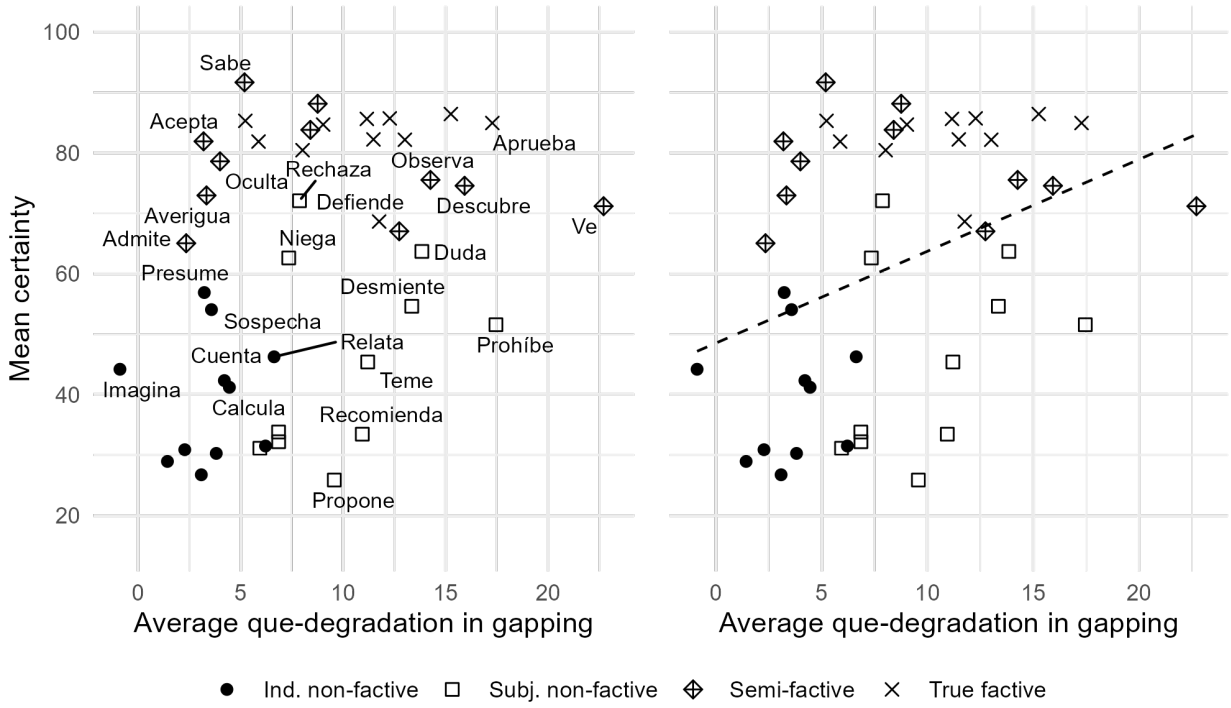


Figure 4.6: Correlation of the mean degradation of gapping with *que* (Chapter 3) and mean certainty rating for the 44 verbs, with labels (left) and with regression line (right).

The low value of R^2 at 0.11 is unexpected. If factivity really is the crucial, determining factor for the acceptability of *que*, this coefficient should be larger, i.e. the correlation should be stronger. Despite the

[14] I explain what a residual is in Section 4.3.1 above.

statistical significance, this weak correlation thus casts doubt on the role of factivity for *que*. Another problem results from the role of one particular verb type, the indicative-taking non-factives (= verbs of saying and thinking). Recall that this verb type was the only one that consistently only showed a negligible degradation of *que* in the experiments in Chapter 3. The low average *que*-effect for this verb type is visible in Figure 4.6, where the black dots representing the verbs of this type all range near the left end of the horizontal scale. This verb type also seems to drive the correlation with the certainty ratings: the black dots are concentrated in an area towards the bottom left in Figure 4.6, which creates the impression that they are "pulling down" on the lower end of the regression line.[15]

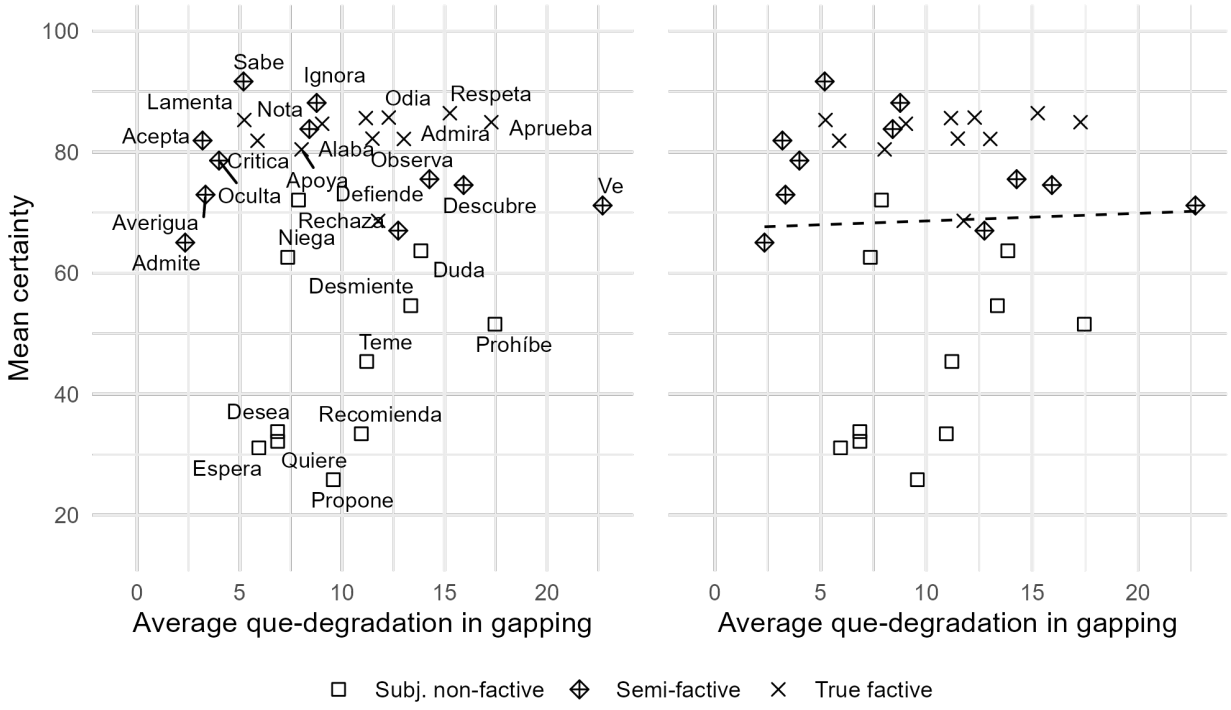


Figure 4.7: Correlation of the mean degradation of gapping with *que* (Chapter 3) and mean certainty rating without the indicative-taking non-factives, with labels (left) and with regression line (right).

To test whether it is indeed the indicative-taking non-factives that drive the correlation, I looked at the data once more, but this time without this verb type. Figure 4.7 shows the resulting correlation of mean certainty rating and average *que*-degradation. This figure shows the same data as Figure 4.6 above, with the difference that the black dots representing the indicative-taking non-factives have been taken out. Without these verbs, the figure still shows a large spread of data points in both the vertical and horizontal directions. This means that the data still includes verbs with high and low certainty ratings (vertical axis), and verbs with large and little *que*-degradation (horizontal axis). If factivity determines the acceptability of *que* in gapping, the correlation of certainty ratings and *que*-degradation that we saw above should thus also surface here. However, the linear model I fitted to these data indicates a much smaller effect ($b = 0.13$)

[15]The negation verbs on the other hand, which I suspected in Section 4.3 might play a special role in the correlation, are all located centrally within the "cloud" of data points and do not seem to have a discernible effect on the slope of the regression line. Unlike for the indicative-taking non-factives, the patterns of the negation verbs thus provide no reason to scrutinize this verb type in particular.

than the model of the data with the verbs of saying and thinking above. The near horizontal regression line in the right side of Figure 4.7 visualizes the small size of the effect. The linear model furthermore indicated that the effect was not statistically significant ($p = 0.87$), and that the correlation coefficient was close to 0 (adjusted $R^2 = -0.03$). Thus, for all intents and purposes, the degradation with *que* does not account for certainty ratings in this sub-set of the data, contrary to expectation.

All in all, the correlations of certainty ratings and *que*-degradation presented this section have provided evidence against the hypothesis that factivity is responsible for the *que*-effect in gapping. This conclusion holds despite the fact that there is a statistically significant, positive correlation of the average certainty rating and *que*-degradation for each verb. The low correlation coefficient, paired with the driving force of the indicative-taking non-factives shows that certainty, and thus by extension factivity, unlikely play a decisive role in the *que*-effect. Instead, the decisive factor is likely an inherent property of the verb type that drives the correlation, i.e. the indicative-taking non-factives. These conclusions are identical to, and thus reinforce, the conclusions I draw in Chapter 3 based on the acceptability data alone.

4.5 Conclusion

In this chapter, I presented the results of a rating study on the interpretation of 44 Spanish clause-embedding verbs. Applying a Spanish equivalent of the *certain that*-diagnostic (Tonhauser et al. 2018, Degen & Tonhauser 2021, 2022), I obtained certainty ratings for each of these verbs. These certainty ratings are indicative of the degree to which these verbs are factive. The results showed that many verbs receive categorically high or low certainty ratings, and may thus, in principle, be classed as factive or non-factive, respectively. However, there were also many verbs that showed intermediate ratings, for which such a categorization was not straightforward. The verbs showed a gradual cline in certainty ratings, which made it impossible to pinpoint a specific cut-off point between factives and non-factives. These results suggest that factivity is a gradual phenomenon that is not ideally captured by the assumption of distinct and mutually exclusive categories like factive and non-factive. This gradualness affected the distinction of semi-factive and true factive even more strongly. In the data there is no evidence for such a split within the factive verbs. In this regard, the results thus parallel Degen & Tonhauser's (2022) results on English, who find a similar gradualness. The seven verbs that feature in both Degen & Tonhauser's (2022) and my study (in their English/Spanish translation equivalents) show comparable certainty ratings.

Despite the gradualness in the certainty ratings, however, my categorization of the verbs in Chapter 3 into factive and non-factive was overall reasonably accurate. Most of the verbs that I had classed as factive received relatively high average certainty ratings, and most of the verbs that I had classed as non-factive received relatively low certainty ratings. This "sanity check" on my judgments of the individual verbs was thus all in all successful. Apart from checking for my sanity, the data collected in the study may prove useful in the investigation of further (linguistic) properties, especially of those that have been claimed to directly connect to factivity. Within the realm of morphosyntax, I discussed two such properties: mood selection and the effect of *que* in embedded gapping. While the subjunctive-taking verbs received statistically significantly higher certainty ratings on average, this effect was so small that it lead me to conclude that the subjunctive is not categorically more factive than the indicative. This finding impacts contributions from the literature that claims such a connection between mood selection and factivity. I specifically

discuss Rivero (1971) in this regard, who claims that verbs with variable mood selection receive factive readings with the indicative, and non-factive readings with the subjunctive. My experimental data on those verbs that Rivero (1971) discusses does not provide any evidence for her claim, since the verbs that she mentions may have high or low certainty ratings with either mood.

In the discussion of *que* in gapping, I found that certainty at first glance appears to have a meaningful connection to the degree of degradation that each verb shows with *que*. There was a statistically significant correlation of the certainty ratings obtained in the experiment of this chapter with the data on *que*-degradation from Chapter 3. However, a more in-depth scrutinizing of this correlation cast doubt on the impact of certainty on *que* in gapping. First, the actual correlation was low, which goes against expectations if factivity determines the acceptability of *que*. Second, the correlation was driven by one particular verb type, the indicative-taking non-factives (i.e. verbs of saying and thinking). Without this verb type, the correlation disappeared. This finding reinforced the conclusion of Chapter 3 that rather than factivity itself, the verbs of saying and thinking play a special role in the acceptability of *que* in gapping (and these verbs happen to tend towards the lower end of the factivity scale). In the following chapter, I try to pin down what this special role of the verbs of saying and thinking is by presenting the results of a study on the meaning of *que* in gapping.

5 An exploratory approach to the meaning of *que* 'that' in Spanish embedded gapping

In this chapter, I approach the meaning contribution of the complementizer *que* 'that' in Spanish globally embedded gapping. Specifically, I investigate whether there is a difference in meaning between gapping with *que* as in (1) and without *que* as in (2).

(1) **Globally embedded gapping with *que*** (from the experimental materials in Chapter 3)
Santiago cree que el rey fuma cada día y **que** la reina, a veces.
Santiago thinks that the king smokes every day and that the queen at times
'Santiago thinks that the king smokes every day and that the queen smokes sometimes.'

(2) **Globally embedded gapping without *que*** (from the experimental materials in Chapter 3)
Santiago cree que el rey fuma cada día y la reina, a veces.
Santiago thinks that the king smokes every day and the queen at times
'Santiago thinks that the king smokes every day and the queen smokes sometimes.'

The motivation for this investigation into the meaning of *que* in gapping stems from the acceptability patterns of gapping with and without *que* that I observed in Chapter 3. In that chapter, the results of a rating study indicate that gapping with *que* may receive degraded acceptability ratings in comparison to the version without *que*, which receives (near) fully acceptable ratings across the board. Whether or not *que* triggers a degradation and the specific degree of the degradation vary depending on the type of embedding verb. In a nutshell, only verbs of saying and thinking (like *cree* 'thinks' in 1 and 2 above) show acceptability ratings with *que* that approach those of the version without *que*. All other verb types that were tested (emotives, knowledge and perception verbs, desideratives, and verbs with implicit negation) show a degradation of *que* to the aforementioned variable degree, which however does not indicate outright unacceptability for any of the verb types. To account for this sensitivity of *que* to the type of embedding verb, I hypothesize in Chapter 3 that *que* in gapping might be associated with a particular meaning, which harmonizes with the meaning of the verbs of saying and thinking (thus leading to no degradation in acceptability), but clashes with the meaning of the other verb types. I speculated that this clash in meaning may be rescued by assigning the particular meaning triggered by *que* to the embedding verb, at the cost of varying degrees of acceptability depending on how easily such a meaning can be constructed for the individual embedding verb. For details on this hypothesized mechanism see Chapter 3.5, where I motivate the hypothesis from a structural point of view.

The obvious question that arose in that discussion is what this "particular meaning" of *que* in gapping might be. As per my speculation, it must be a meaning component characteristic of the verbs of saying and thinking, so that it does not lead to the degradation in acceptability in this type of verb that I observed for all other types of verbs. In short, the discussion in Chapter 3.5 revealed four candidates for this

meaning component: factivity, assertion, at-issueness, and reportiveness. The present chapter explores these four aspects of meaning in their connection to *que* in gapping. To do so, the present chapter tests the gapping materials of the acceptability experiments in Chapter 3 once more. Specifically, I conducted two rating experiments to answer the question whether *que* in gapping is associated with any of the four aforementioned meaning components.

For the first of the four meaning candidates, factivity, I have shown already at some length that it is unlikely to be responsible for the *que*-effect in gapping: in Chapter 4, I provided experimental evidence that factivity does not seem to have a meaningful relation to the degree to which a verb penalizes *que* in gapping. The measure of this relation in Chapter 4 was indirect: I compared pooled data from the acceptability experiments on gapping from Chapter 3 to factivity data for the individual verbs that was independently obtained by means of a different experiment. In the present chapter, I return to the issue of factivity once more by means of a more direct approach to factivity in gapping: Experiment 1 applied the diagnostic for factivity from Chapter 4 (the *certain that*-diagnostic, see Chapter 4.1) directly to gapping materials from Chapter 3 such as (1) and (2) above. The results of Experiment 1 will show no discernible effect of *que* on certainty ratings, suggesting once more that speaker certainty (and factivity) have little to do with the acceptability *que* in gapping.

In Experiment 2, I explore the remaining three meaning candidates (i.e. assertion, at-issueness, and reportiveness), by applying individual diagnostics for these three properties to gapping with and without *que* (i.e. the same materials as for Experiment 1). The diagnostics used in Experiment 2 constitute innovative means to obtain semantic/pragmatic rating measures for assertion, at-issueness, and reportiveness. To my knowledge, there are no other contributions in the literature that apply these exact diagnostics to gain data on the properties in question. This means that while Experiment 2 was confirmatory in design, it was to a considerable degree also exploratory "in spirit" due to the innovative nature of the diagnostics. The discussion of the diagnostics will reveal that they did not ideally capture the three properties they were meant to measure, which provides some problems in the interpretation of the results. I still chose to include Experiment 2 in this dissertation because from the "big picture perspective" it still provides relevant information. It still targets the relevant overarching issue of the meaning of *que* in gapping, which is given independently of whether the diagnostics ideally capture the specific properties assertion, at-issueness or reportiveness. However, this information is unfortunately entirely negative: the results of Experiment 2 show no discernible effect of the presence or absence of *que* on interpretation as far as these properties are concerned. In this regard, the results of Experiment 2 are similar to the results of Experiment 1.

In sum, the results of the two experiments in this chapter provide no evidence that *que* in gapping has the specific effects on interpretation that I hypothesized in Chapter 3.5, or any such effect for that matter. As far as the investigation of *que* in gapping is concerned, the present chapter thus takes us no further towards an explanation of the acceptability patterns of *que* with the different verb types (which is the opposite of what I initially hoped it would do). The only potential positive conclusion that I see from the results in this chapter is that I did my due diligence in testing the hypotheses on the effect of *que* on interpretation I made in Chapter 3, even though these hypotheses must be rejected in light of the results. The structure of this chapter is straightforward: Section 5.1 contains Experiment 1, Section 5.2 contains Experiment 2, including a description and critical discussion of the innovative diagnostics, and Section

5.3 repeats the conclusions that I have just made regarding the overall impact of the two experiments on this dissertation.

5.1 Experiment 1: Certainty

Experiment 1 applied the *certain that*-diagnostic for factivity (Chapter 4.1) to the experimental materials of Chapter 3.[1] The reasoning for testing the gapping materials with the *certain that*-diagnostic was that if *que* in gapping causes a non-factive interpretation, the utterer of the gapping structure should be interpreted to not presuppose the propositional content of the complement clause (= the 'CC') in gapping with *que*, irrespective of the (type of) embedding verb. This reasoning predicts that certainty ratings should be low across the board for gapping with *que*. In gapping without *que*, presupposition and (thus certainty) of the CC should 'proceed as normal', i.e. as a function of the embedding verb: with factives the CC should be presupposed (= high certainty ratings), with non-factives the CC should not be presupposed (= low certainty ratings). Given that the formulation of my hypothesis is strong (*que* outright causes a non-factive interpretation, and does not just favour or enable it), this predicts a strong effect of *que* rather than a small or gradual effect.

Experiment 1 had a 1×2 design, the sole factor being the presence vs. absence of *que*. For a sample item in both conditions, see (3). This item is nearly identical to an item from Experiment 1 in Chapter 3, with the crucial difference that it is presented in a direct speech context. This adjustment of the items was necessary since the *certain that*-diagnostic measures the certainty of the speaker (*María* in 3). The embedded gapping configuration of every item thus came as the direct speech complement to the speaker, which was invariably *María* in all items. The items were preceded by 1-2 sentences of context, which had already been a feature of the experimental materials in Chapter 3.2, see there for more details. This context was characterized as the conversational context in the experimental instructions for the participants (rather than the situation participants were asked to imagine being in, see below). Participants gave their certainty ratings by answering a question like (4), which assessed the speakers certainty of part of the CC, i.e. the proposition encoded in the elliptical conjunct.

(3) **Sample item Experiment 1, conditions with and without *que***

María: Santiago cree que el rey fuma cada día y {**que** / Ø} la reina, a veces.
María Santiago thinks that the king smokes every day and that the queen at times
'María: Santiago thinks that the king smokes every day and that the queen smokes sometimes.'

(4) **Diagnostic question Experiment 1**

¿María está segura de que la reina fuma a veces?
María is certain of that the queen smokes at times
'Is María certain that the queen smokes sometimes?'[2]

[1] Specifically, the materials were the items with gapping from Experiments 1, 2, and 3-6 in Chapter 3. These experiments tested intransitive complement clauses. The materials of Experiment 7, which tested transitive clauses, did not feature in the two experiments in this chapter. Experiment 4 tested non-elliptical clauses, which are entirely irrelevant for this chapter.

[2] Notice that this question has declarative word order (i.e. SVO). This word order is generally possible for questions in Spanish and does not seem to carry special requirements on context or trigger strongly different interpretations from questions with subject-verb inversion. The diagnostic question in the experiment in Chapter 4 was presented to participants with subject-verb inversion (*Está María segura de que…*), i.e. in a different word order from this question. This difference between

As mentioned above, the experimental materials of Experiment 1 were identical to a subset of the materials in Chapter 3. They contained the 48 embedding verbs in Table 5.1, and 24 embedded clauses (one of which was *el rey fuma cada día y la reina, a veces* in example 3 above). Each verb combined with two embedding clauses to form a total of 96 sentences like (3), which in the two conditions made a total of 192 test items (= two for each of the 48 embedding verbs, see Chapter 3.2 for details on the design). These items were distributed across eight lists in a Latin square design. Each list contained 24 test items so that each participant judged each embedded clause once and half of the embedding verbs. The distribution of the four verb types I have assumed throughout this dissertation (see Table 5.1) was balanced on each list. Additionally to the test items, every list featured ten declarative control items and ten question control items (see below), for a total of 44 sentences to be judged on each list.

Table 5.1: Verbs of the experiments by their types (in the third person singular present indicative, like in the experimental items).

Non-Factives (with indicative)		**True Factives** (with subjunctive)		**Non-Factives** (with subjunctive)		**Semi-Factives** (with indicative)	
Spanish	*English*	*Spanish*	*English*	*Spanish*	*English*	*Spanish*	*English*
afirma	claims	admira	admires	demanda	demands	acepta	accepts
calcula	estimates	alaba	praises	desea	wishes	admite	admits
considera	considers	apoya	supports	desmiente	denies	averigua	finds out
cree	thinks	aprueba	approves	duda	doubts	conoce	knows
cuenta	tells	consiente	tolerates	espera	expects	descubre	discovers
dice	says	critica	criticizes	niega	denies	ignora	ignores
imagina	imagines	defiende	defends	prohíbe	forbids	nota	notes
piensa	thinks	desaprueba	disapproves	propone	proposes	observa	observes
presume	presumes	detesta	detests	quiere	wants	oculta	hides
relata	relates	lamenta	regrets	rechaza	rejects	oye	hears
sospecha	suspects	odia	hates	recomienda	recommends	sabe	knows
supone	supposes	respeta	respects	teme	fears	ve	sees

One major difference between Experiment 1 and the other certainty experiment in this dissertation, i.e. the experiment in Chapter 4, is that Experiment 1 involved declarative sentences, as opposed to the polar questions in Chapter 4. I chose not to change the existing materials into polar questions for a number of reasons: first, the items already contained considerable syntactic and semantic complexity (clausal coordination, clausal subordination, and ellipsis), which would have been added to by question formation. In order to avoid confusing participants with resulting sentences, I chose to proceed with declaratives. Second, the relevant point of comparison to the data of Experiment 1 are the data obtained in Chapter 3, which also tested declaratives. Since my goal is to account for the acceptability patterns I observed in those experiments, the materials should remain as constant as possible to avoid the introduction of further variation. Third, while the necessary changes to the items themselves (i.e. forming a question from a declarative) would have been straightforward, the presentation of the thus changed items in their context would not have possibly lead to difficulties, since the contexts were constructed to provide a frame for the use of declaratives and not questions.

experiments was an oversight; however I do not assume that this difference had any effect on certainty ratings. (Compare also the discussion of the choice of mood in the diagnostic question in Chapter 4.2, which I similarly do not assume affect certainty ratings.)

However, the use of the *certain that*-diagnostic with declarative clauses leads to conceptual problems regarding the way the resulting ratings relate to factivity. Recall that one hallmark of factive presupposition (as for any presupposition) is its projection out of entailment cancelling environments (see Chapter 3.1.1). Providing such an entailment cancelling environment was the reason to present the items in the experiment in Chapter 4 as polar questions in the first place. The declarative clauses in Experiment 1, like the one in (3) above, feature no such entailment cancelling environment. Thus, in Experiment 1, I cannot equate the certainty ratings in a straightforward fashion to factivity, like I did with the ratings for the polar questions in Chapter 4.

To illustrate this problem, consider the sentence in (5), which was one of the declarative filler items in the experiment. When targeted by the *certain that*-diagnostic as in (6), the propositional content of the declarative clause in (5) should receive very high certainty ratings, since the speaker (*María*) should be interpreted to be certain that the propositional content of the assertion that she is making in (5) is true. However, the expected high certainty ratings of course are not indicative of presupposition: the speaker does not *presuppose* the content of the declarative sentence despite her high certainty of that content. The resulting difficulty in interpreting the certainty ratings for the declarative items would have provided a considerable conceptual problem for me to resolve,[3] had Experiment 1 returned any non-null results. Given that Experiment 1 only returned null results, I will not undertake any in-depth attempt at resolving this conceptual problem, since the results provide no insight on any specific meaning contribution of *que* in gapping anyway. I will, however, return to the question of how the ratings relate to factivity in the case of the experimental items (which involved clausal embedding) further below in the discussion of the results.

(5) **Declarative filler, Experiment 1**

María: Julio adora el tiempo soleado y fantástico de Gran Canaria.
María Julio loves the weather sunny and fantastic of Gran Canaria
'María: Julio loves the sunny and fantastic weather of Gran Canaria.'

(6) **Diagnostic question to the declarative filler, Experiment 1**

¿María está segura de que Julio adora el tiempo soleado y fantástico de Gran Canaria?
María is certain of that Julio loves the weather sunny and fantastic of Gran Canaria
'Is María certain that Julio loves the sunny and fantastic weather of Gran Canaria?'

The certainty ratings for the declarative fillers like the one in (5) above provided a baseline for certainty. The materials also featured polar questions as fillers, the ratings for which provided a baseline for non-certainty. The declarative fillers had been part of the materials in Chapter 3; the question fillers had served as fillers for the experiment in Chapter 4. For an example of a question filler, see the sentence in (7) and its diagnostic question in (8).

(7) **Question filler, Experiment 1**

María: ¿Elena estudia toda la noche para un examen?
María Elena studies all the night for an exam
'María: Does Elena spend all night studying for an exam?'

[3] Possibly, the certainty ratings of Experiment 1 are indicative of veridicality (see Chapter 3.1.1) instead of factivity. Since the experimental materials contained no veridical verbs that were not also factive, the certainty ratings of the individual verbs will not shed light on the question of whether certainty equates to factivity or veridicality (or neither) in Experiment 1.

(8) **Diagnostic question to the question filler, Experiment 1**

¿María está segura de que Elena estudia toda la noche para un examen??
María is certain of that Elena studies all the night for an exam
'Is María certain that Elena spends all night studying for an exam?'

The experimental set-up was identical to that of the experiment in Chapter 4. Participants were asked to imagine that they were attending a party hosted by their friend María. María was described as a chatty person who loved talking about other people. Upon entering the kitchen, participants were asked to imagine overhearing a statement (= the experimental items and declarative fillers) or question (= the question fillers) that María made, within the conversational context provided.[4] Participants then indicated how certain they thought María was regarding the target contents (= the CC for items, the whole content of the utterance for the two types of fillers). To do so they used a slider on a continuous scale with the extremes *sí* 'yes' and *no* 'no', like in the experiment in Chapter 4. Responses were coded as integers from 1 to 100.

The experiment was an online questionnaire hosted on SoSciSurvey.de (Leiner 2024). I used Prolific.com to recruit participants. To ensure the quality of the data, the questionnaire link was available only to those 238 people that had participated in the experiment in Chapter 4 and that were not excluded by the exclusion criteria of that experiment.[5] These were monolingual native speakers of European Spanish who grew up speaking their native language only. In total, 80 participants completed the questionnaire. I set an exclusion criterion based on the certainty ratings for the controls: only the data from those participants whose ratings indicated a statistically significant difference between declarative and question controls (as determined by a *t*-test for each participant) entered the analysis. All participants passed this criterion, i.e. they all made the expected difference between declarative and question fillers. Consequently the data of all 80 participants entered the analysis.

The results are in Table 5.2 and Figure 5.1. The ratings for the filler controls show that the *certain that*-diagnostic worked as expected: declarative fillers received high certainty ratings, and question fillers received low certainty ratings. The ratings for the two experimental conditions, *que* vs. no *que*, show hardly any discernible difference, let alone the strong effect predicted from my hypothesis (see above).[6] I conclude that *que* in gapping has no effect on meaning that relates to the interpretation of speaker certainty in any way, whether this extends to factivity or not.

A closer look at the four individual verb types reinforces this conclusion. As can be seen in Table 5.3 and Figure 5.2, the effect of *que* on certainty ratings within the verb types is just as negligible as for the overall data. Overall, the two factive verb types received higher certainty ratings than the two non-factive

[4] I acknowledge that it may be a difficult situation to imagine being in where you accidentally overhear the utterance someone has made, but where you are simultaneously aware of the conversational context of that utterance. I proceed under the assumption that the participants were not so confused by this slight incoherence in the experimental setup that it affected their ratings tremendously. A further issue with the experimental setup were the varying degrees of degradation of the items in the condition with *que*. These degraded items may have been difficult for participants to interpret regarding speaker certainty. To go some way in alleviating this difficulty, the instructions made participants aware that not all of María's utterances may sound very natural, and that they should try to give an intuitive judgment as well as they could.

[5] Experiment 1 ran six weeks after Block 2 of the experiment in Chapter 4. It is therefore unlikely that participants remembered the details of their preceding experiment.

[6] Since the size of the effect (or rather lack thereof) sufficiently illuminates the role of *que* in the interpretation of speaker certainty, I did not perform an inferential analysis on the data. I do not believe the results of such an analysis would have contributed anything to the discussion.

Table 5.2: Experiment 1 (certainty ratings): Mean ratings per experimental condition and control items (*N* = number of observations, *SD* = standard deviation, *SE* = standard error, *CI* = 95% confidence interval)

Complementizer	N	Mean rating	SD	SE	CI
Ø	960	57.22	43.86	1.42	2.78
que	960	58.02	42.28	1.36	2.68
Declarative controls	800	94.61	15.39	0.54	1.07
Question controls	800	18.46	27.90	0.99	1.94

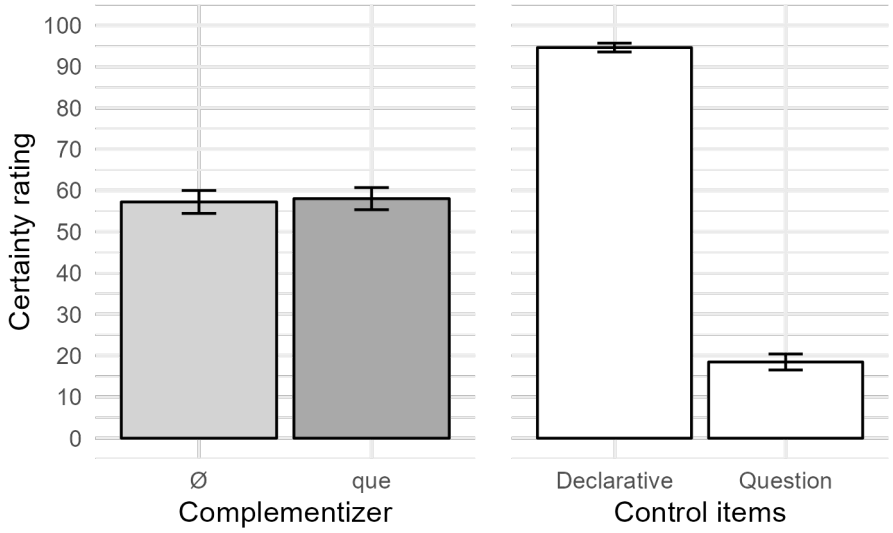


Figure 5.1: Mean ratings and 95% confidence intervals for experimental conditions and controls (Experiment 1).

types, which suggests that the *certain that*-diagnostic may have picked up on factivity in this experiment after all, at least in the experimental items with clausal embedding, where the diagnostic targeted the CC. The mean ratings for the factive and non-factive types are in the same ballpark as their mean certainty ratings in the experiment in Chapter 4.3. For mood however, the results of Experiment 1 seem to reverse the results of the experiment in Chapter 4.3: for both factive and non-factive verbs, the indicative-taking ones received slightly higher mean certainty ratings that the subjunctive-taking ones. In the experiment in Chapter 4, the opposite was the case: the subjunctive-taking verbs in both the factive and non-factive groups received higher certainty ratings.

To further investigate the effect of mood on certainty ratings in Experiment 1, I used `R` (R Core Team 2023) and the `afex`-package (Singmann et al. 2022) to fit a linear mixed effects model on the certainty data. The model contained the fixed factors mood (indicative vs. subjunctive) and factivity (factive vs. non-factive), as well as the interaction of the fixed factors. The random structure was the maximal possible randon structure: random intercepts and slopes for the 24 embedded contents (≈ items) and participants. Fixed factors were sum-to-zero contrast coded to facilitate the interpretation of interaction effects. The model indicated that the effect of mood that I observed above was statistically significant ($b = -3.31, SE = 1.33, t = -2.48, p = 0.02$). As one would expect, the effect of factivity was significant as well ($b = 15.15, SE = 1.18, t = 12.89, p < 0.001$). The interaction effect was not significant ($b = 0.38, SE = 1.19, t = 0.32, p = 0.75$).

Table 5.3: Experiment 1 (certainty ratings): Mean ratings per verb type and experimental condition, and control items (*N* = number of observations, *SD* = standard deviation, *SE* = standard error, *CI* = 95% confidence interval)

Factivity	Mood	Complementizer	N	Mean rating	SD	SE	CI
Non-factive	Indicative	Ø	240	46.92	29.31	1.89	3.73
Non-factive	Indicative	*que*	240	45.70	29.08	1.88	3.70
Non-factive	Subjunctive	Ø	240	38.48	30.24	1.95	3.84
Non-factive	Subjunctive	*que*	240	38.90	28.90	1.87	3.67
Semi-factive	Indicative	Ø	240	74.13	27.84	1.80	3.54
Semi-factive	Indicative	*que*	240	77.08	24.54	1.58	3.12
True factive	Subjunctive	Ø	240	69.35	28.99	1.87	3.69
True factive	Subjunctive	*que*	240	70.40	25.08	1.62	3.19

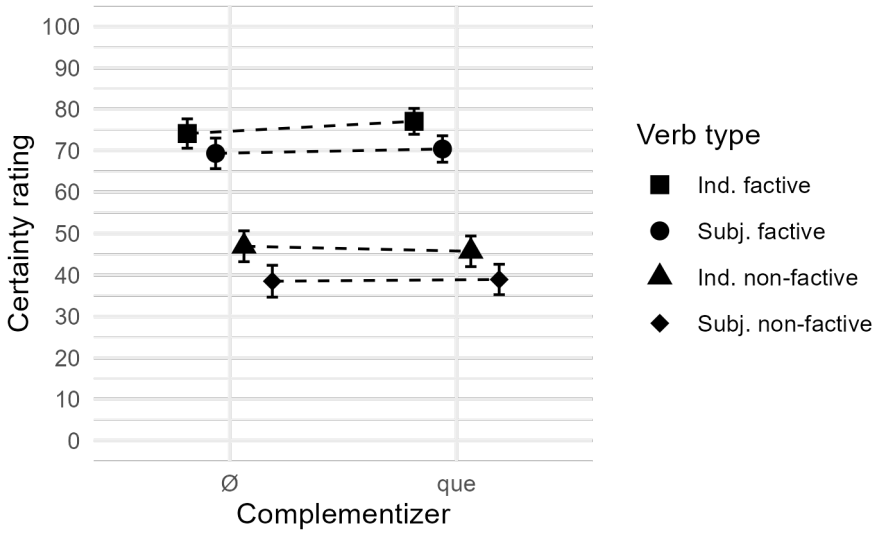


Figure 5.2: Mean ratings and 95% confidence intervals for verb types by experimental conditions and controls (Experiment 1).

The significant effect of mood selection suggests that the higher certainty ratings for the indicative-taking verbs in Experiment 1 are indicative of a genuine pattern. An exhaustive investigation of this pattern, especially under consideration of the differences between Experiment 1 and the experiment in Chapter 4 is beyond the scope of the present chapter. I will briefly touch upon three possible reasons that I see for why the indicative-taking verbs might have received higher certainty ratings in Experiment 1: one, Experiment 1 contained four more verbs (one of each type) than the experiment in Chapter 4. These verbs were: *considera* 'considers' (indicative-taking non-factive), *demanda* 'demands' (subjunctive-taking non-factive), *conoce* 'knows' (semi-factive), and *consiente* 'tolerates' (true factive). It is possible that these four verbs play crucial role in the mood effect. However I have no reason to suspect that these four verbs specifically caused mood and certainty to interact to such an extent that it reversed the effect observed in Chapter 4. But it is possible nevertheless.

Two, the effects of the individual items in Experiment 1 were less controlled for than in the experiment in Chapter 4. In Chapter 4, every embedding clause combined with every embedding verb, which allowed for an interpretation of the results which abstracted away from the specific CCs. As has been shown in previous research, properties of the CC such as plausibility may affect certainty ratings (Degen &

Tonhauser 2021). Since in Experiment 1 every verb only combined with two out of the 24 embedded clauses, such an abstraction from the CCs is not justified for Experiment 1. A discussion of the ratings for the individual complement clauses would take us too far beyond the original point of this chapter, the meaning of *que* in gapping. However, it cannot be ruled out that what seems to be a mood effect in Figure 5.2 is actually indicative of an item effect.

Three, the difference may be triggered by a semantic or pragmatic difference between declarative clauses (Experiment 1) and polar questions (experiment in Chapter 4). Of course, the coordinate embedded clauses involving gapping of Experiment 1 are hardly representative of declarative clauses as such and do not provide an ideal basis of comparison to the simplex embedded clauses to the test sentences of the experiment in Chapter 4. More research is thus needed on (all three of these options but especially) this third option.

What remains as the main result of Experiment 1 is that *que* in gapping does not seem to affect the interpretation of speaker certainty in any way. The materials of Experiment 1 contained declarative clauses, which means that the certainty ratings of Experiment 1 cannot be straightforwardly interpreted as being indicative of factivity. However, if *que* – as per my hypothesis – determines a non-factive interpretation, one may still confidently expect certainty ratings to be low across the board for gapping with *que*. This expectation is not met in the data. I therefore conclude that *que* does not affect the interpretation of factivity.

5.2 Experiment 2: Assertion, at-issueness, and reportiveness

Experiment 2 constitutes an attempt at approaching the way *que* in gapping (potentially) modulates three further semantic/pragmatic properties: the interpretation of assertion, at-issueness, and reportiveness. As far as the gapping sentences are concerned, the experimental materials of Experiment 2 were identical to those of Experiment 1. The control items were different from those in Experiment 1, I will return to them below. A more crucial difference between the experiments lay with the diagnostics I applied to the test items. Thus, like in Experiment 1, the test sentences came in two conditions, i.e. with *que* and without *que*, see (9). For a description of the number of embedding verbs, embedded clauses, total items, and the division of these items onto lists, see Section 5.1.

(9) **Sample item Experiment 2, conditions with and without *que*** (identical to the sample item of Experiment 1 in example 3 above)
María: Santiago cree que el rey fuma cada día y {**que** / Ø} la reina, a veces.
María Santiago thinks that the king smokes every day and that the queen at times
'María: Santiago thinks that the king smokes every day and that the queen smokes sometimes.'

For each of the experimental items, participants indicated whether they agreed or disagreed with three statements. For this, participants clicked one of two boxes per statement, which were labelled *sí* 'yes' and *no* no.[7] These statements functioned as the diagnostics for the three semantic/pragmatic properties

[7]Since the number of judgments increased from one in Experiment 1 to three in Experiment 2, I did not implement the slide rule on the continuous scale for Experiment 2, but instead opted for categorical *yes/no*-options. I did this in the hope that it would enable participants to give judgments more quickly and thus put less strain on them. Given the "exploratory spirit" of Experiment 2, this procedure seemed justified.

that the experiment was supposed to test. For the example item in (9) above, participants gave judgments on the three statements in (10) (for assertion), (11) (for at-issueness), and (12) (for reportiveness). For all three of these diagnostics, the predictions are as follows: if *que* determines interpretation along these lines (i.e. an assertive, at-issue and/or reportive interpretation), the overall proportion of *yes* answers in the condition with *que* should be high, and the *yes*-proportion should also be noticeably higher than in the condition without *que* overall. Furthermore, since the indicative-taking non-factives (per assumption) already carry an assertive, at-issue and/or reportive meaning, they should display high *yes*-proportions in both conditions. For the verbs of this type particularly, there should be hardly any effect of *que*, while the other three verb types should show an effect of *que*. Since my hypothesis is that *que* determines interpretation (rather than only enable or favour a specific interpretation), we expect this effect to be large for the three respective verbs types (i.e. the subjunctive-taking non-factives, semi-factives, and true factives).

(10) ***Possible that not*-diagnostic (for assertion)**
Es posible que la reina no fume a veces.
is possible that the queen NEG smokes-SUBJ at times
'It is possible that the queen does not smoke sometimes.'

(11) **Importance diagnostic (for at-issueness)**
Lo más importante que dice María es que la reina fuma a veces.
the most important that says María is that the queen smokes-IND at times
'The most important thing that María is saying is that the queen smokes sometimes.'

(12) **Reportiveness diagnostic**
Santiago le contó algo nuevo a María, concretamente que la reina fumaba a veces.
Santiago her told something new to María specifically that the queen smoked-IND at times
'Santiago told María something new, specifically that the queen smokes at times.'

The diagnostic for assertion in (10) asked participants whether it was possible that the CC in the second, i.e. elliptical, conjunct was false. This diagnostic relates to one of the characteristics of asserted content, which is that asserted content can be easily picked up by interlocutors in discourse for approval or rejection. This is illustrated in (13), which builds on a simplified, English version of the experimental item in (9) above. If we assume that the speaker in (13) *A* asserts the CC of the embedding verb *thinks*, their interlocutor *B* should easily be able to deny the CC. This ability to be targeted for denial in discourse contrasts asserted content from non-asserted content, such as the content of a nominal apposition. An example of a nominal apposition is in (14), where the DP *an avid smoker* is an apposition to the DP *the queen*. In a sentence like that of speaker *A* in (14), only the content of the clause without the apposition should be a potential target for denial in discourse, since it is this content that is asserted. Hence, the utterance of interlocutor *B*, who denies the (asserted) content of the clause, is felicitous. The utterance of interlocutor *B'* on the other hand is not felicitous, since it targets the content of the apposition for denial, which is not asserted.

(13) A: Santiago thinks that the queen smokes sometimes.
B: No, that is not true, the queen doesn't smoke at all.

(14) A: The queen, an avid smoker, has gone to Cuba on holiday.
B: No, that is not true, the queen hasn't gone to Cuba.
B': #No, that is not true, the queen doesn't smoke at all.

The *possible that not*-diagnostic approaches this possibility of denial of the CC in discourse indirectly by asking whether the CC is potentially false. The reason for this indirect approach lies with the fact that the experimental design did not straightforwardly allow for an implementation of the necessary discourse aspect to test denial directly. This is to say that the experimental materials simply did not come in dialogue form. I acknowledge that asking whether the CC is possibly false does not closely correspond to asking whether the CC can be denied in discourse. The diagnostic entails only logical-semantic properties (i.e. truth vs. falsity) and does not relate to discourse properties at all. The filler items in Experiment 2, which amongst other discourse features contained nominal appositions like in (14) above, go some way in checking whether the diagnostic picked up on assertion. However, the tension between what the diagnostic was supposed to measure and (the open question of) what it actually did measure still remains. I propose that this tension may be tolerated as part of the "exploratory spirit" of Experiment 2. Like for the diagnostic in Experiment 1, this tension would have provided a considerable conceptual problem for me to resolve, had there been any non-null results to Experiment 2. Since Experiment 2 provided null results, I will not further dwell on what exactly this diagnostic tells us.

The other two diagnostics capture the properties they are intended to measure in a more straightforward manner. The diagnostic for at-issueness in (11) above asks participants whether the CC is the most important part of the speaker's utterance. This formulation of the diagnostic relates to the intuitive formulation that the at-issue content of an utterance is the "main point" of that utterance (Simons et al. 2010). Asking for the main point by way of "the most important thing" provides a straightforward, intuitive way for participants to indicate their judgments of the at-issueness of the CCs.

The diagnostic for reportiveness in (12) above aimed to measure whether the embedding verb in the experimental items received a verb-of-saying interpretation. Under this interpretation, the sentence in the sample item in (9) above, reproduced below in (15), would constitute a report on what the embedding subject *Santiago* has told to the speech-act subject *María*. The diagnostic sentence for this item in (16), repeated from (12) above, targets this interpretation by asking whether the embedding subject has told the CC to the speech-act subject.[8]

(15) **Sample item, Experiment 2** (same as in example 9 above)
María: Santiago cree que el rey fuma cada día y {**que** / Ø} la reina, a veces.
María Santiago thinks that the king smokes every day and that the queen at times
'María: Santiago thinks that the king smokes every day and that the queen smokes sometimes.'

[8]The reportiveness diagnostic features the phrase *algo nuevo* 'something new' as the direct object to *contó* 'told'. The actual content of the complement clause, which is crucial for the diagnostic, is presented as a clausal apposition to this direct object. This feature of the diagnostic is unnecessary, a direct combination of the embedding verb with the crucial embedded clause would have been just as appropriate to measure reportiveness. In fact, the filler items (see Section 5.2.2) featured such a direct embedding for the reportiveness diagnostic. The introduction of the *something new* phrase for the experimental items was an oversight. I proceed under the assumption that this phrase had no influence on responses.

(16) **Reportiveness diagnostic**
Santiago le contó algo nuevo a María, concretamente que la reina fumaba a
Santiago her told something new to María specifically that the queen smoked-IND at
veces.
times
'Santiago told María something new, specifically that the queen smokes at times.'

As can bee seen in (15) and (16), the tenses in the two sentences are non-identical. In the target sentence (15), the embedding predicate *cree* 'thinks' is in the present tense (as is the predicate in the embedded clause, *fuma* 'smokes'), while in the diagnostic sentence, the matrix verb *contó* 'told' is in the past. This difference in tense is necessary because for María to report on something another person has told her, that telling event must have necessarily taken place in the past. The question that arises from this is whether it would not have been more natural for the embedding verb in the item to be in the past as well, rather than in the present. The reason for implementing these verbs in the present is that I want to account for the patterns observed in the acceptability study in Chapter 3, where the embedding verbs were in the present as well. The resulting soft clash in tense between the target sentence (present) and the diagnostic (past) may, once more, be tolerated under consideration of the "exploratory spirit" of Experiment 2. As for all other diagnostics discussed in this Chapter, i.e. of both experiments, a follow-up experiment would have been appropriate that employed a refined way of measurement, had the results been such that they called for such a follow-up. Due to the null results of Experiment 2, no further experiment with such a refined way of measurement was necessary.

In sum, the three diagnostics in (10) to (12) approach ways to measure the three crucial discourse properties relevant to Experiment, i.e. assertion, at-issueness, and reportiveness. As I discuss openly above, these measurements may not ideally capture these three properties, especially that for assertion provides some conceptual problems. I refer once more to the "exploratory spirit" of Experiment 2, which justifies the loose approximation of the diagnostics to the properties in question. However, as for the measurement in Experiment 1, I believe that such a loose approximation would have been enough to return any indication of the effect of *que* on interpretation, which as per my hypothesis should have been large. In other words, I deem that my diagnostics (with all the flaws they may have) would have been sensitive enough to pick up on an effect of *que* on interpretation, had there been such an effect. I will return to the diagnostics in Section 5.2.2, where I scrutinize them more closely by means of the experimental results for the filler items. First, I discuss the experimental results (Section 5.2.1). I chose this slightly backward procedure of reporting the results before more closely examining the diagnostics because the examination of the diagnostics relies on the experimental results, and because the experimental results were null results (as mentioned above), which means they can be dealt with relatively quickly.

The overall procedure for Experiment 2 was like for Experiment 1. It was an online questionnaire hosted on SoSciSurvey.de, with participant recruitment via Prolific.com. 80 native speakers of European Spanish completed the questionnaire. Four of these participants failed the attention checks (see Section 5.2.2), so that the data of 76 of them entered analysis. The experimental items were distributed across eight lists in a Latin square design, just like in Experiment 1. In addition to 24 experimental items, each list contained 16 filler items to validate the diagnostics (see Section 5.2.2). Thus, each participants gave three judgments for 40 sentences. Due to an oversight in the maintenance of the online questionnaire tool, the first 54

participants were not forced to give judgments. This means participants could go straight to the next page of the questionnaire without clicking *sí* or *no* for any of the diagnostic statements. This oversight was amended after participant 54, such that the remaining 26 participants had to give judgments for all diagnostics to proceed in the questionnaire. Due to this oversight, a total of 20 items and 3 fillers did not receive an answer for one of their three statements.

5.2.1 Experimental results

The results for the three diagnostics are in Figures 5.3 (assertion), 5.4 (at-issueness) and 5.5 (reportiveness). As can be seen in these figures, the effects of *que* on the interpretation of these properties is very small across the board. For all three properties, these small effects obtain for the total data including all verbs (left sides of the figures), as well as for every of the four individual verb types (right sides of the figures). Even for those effects that seem to be stronger than others (e.g. the at-issueness ratings for the indicative-taking non-factives in Figuer 5.4), the error bars indicating the 95% confidence intervals overlap by a considerable margin. I conclude that *que* in gapping does not have the strong contribution to meaning that I hypothesized in Chapter 3. Since a visual inspection of the figures provides sufficient evidence for this conclusion, I will not provide any inferential statistics for the data of Experiment 2. I do not believe that the discussion of *p* values, had they indicated statistical non-significance or (against expectations) statistical significance, would contribute anything to our understanding of the meaning of *que* in gapping. In the rest of this section, I will not further discuss the (non-)effect of *que*, and instead focus on the overall patterns across the four verb types for a brief description of the three diagnostics.

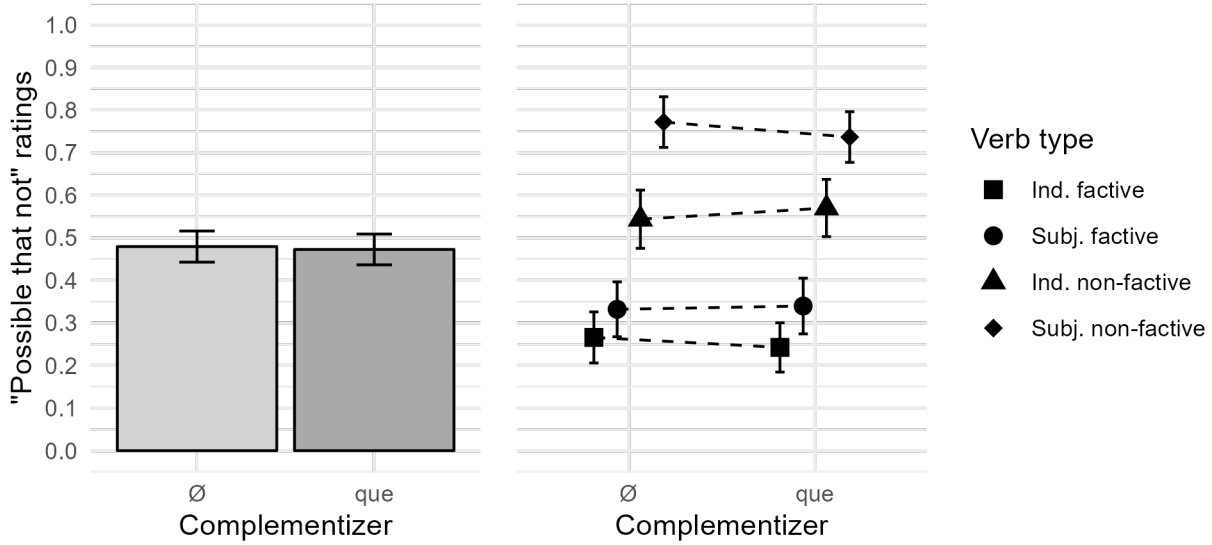


Figure 5.3: *Possible that not* ratings for assertion: Proportion of *sí* "yes"-responses and 95% confidence interval by experimental condition (Experiment 2). For all embedding verbs (left) and for the four verb types (right).

The *possible that not*-diagnostic was supposed to measure assertion. Recall that, according to my assumptions on the four verb types in Chapter 3.1.1, only the indicative-taking non-factive verbs may embed assertions. The two factive verb types, as well as the subjunctive-taking non-factives do not embed asser-

tions, according to these assumptions. Figure 5.3 shows that, for the two factive types and the indicative-taking non-factives, the *possible that not*-diagnostic overall returned proportions of *yes*-responses that are in line with these assumptions: the two factive types receive a relatively low *yes*-proportion, with a negligible difference between them. The indicative-taking non-factives receive a notably higher proportion of *yes*-responses. However, the fourth verb type, the subjunctive-taking non-factives, do not fit the expected pattern: they receive the highest proportion of *yes*-responses out of the four verb types. The *possible that not*-diagnostic thus likely did not actually measure assertion. I will return to a further attempt at validating this diagnostic in Section 5.2.2.

The relatively high proportion of *yes*-responses for the subjunctive-taking non-factives might instead be indicative of a negative inference these verbs carry with regard to their CC. This negative inference might be identical to what Meinunger (2017) calls "counter-factivity" (which I touch upon briefly in Chapter 3.1.1). Whether it is indeed a presupposition that the CC is false as suggested by Meinunger's (2017) term "counter-factivity", or whether this inference only holds in declarative clauses is an issue for further research. Recall that in the experiment in Chapter 4, one sub-type of the indicative-taking non-factives, verbs with implicit negation, actually received relatively high certainty ratings in the polar questions that that experiment tested. These high certainty ratings in polar questions suggest that the negative inference may not be constant in entailment cancelling environments, which would mean that it does not constitute a presupposition. Again, more research is needed on the inferences triggered by the subjunctive-taking non-factives in different sentence environments.

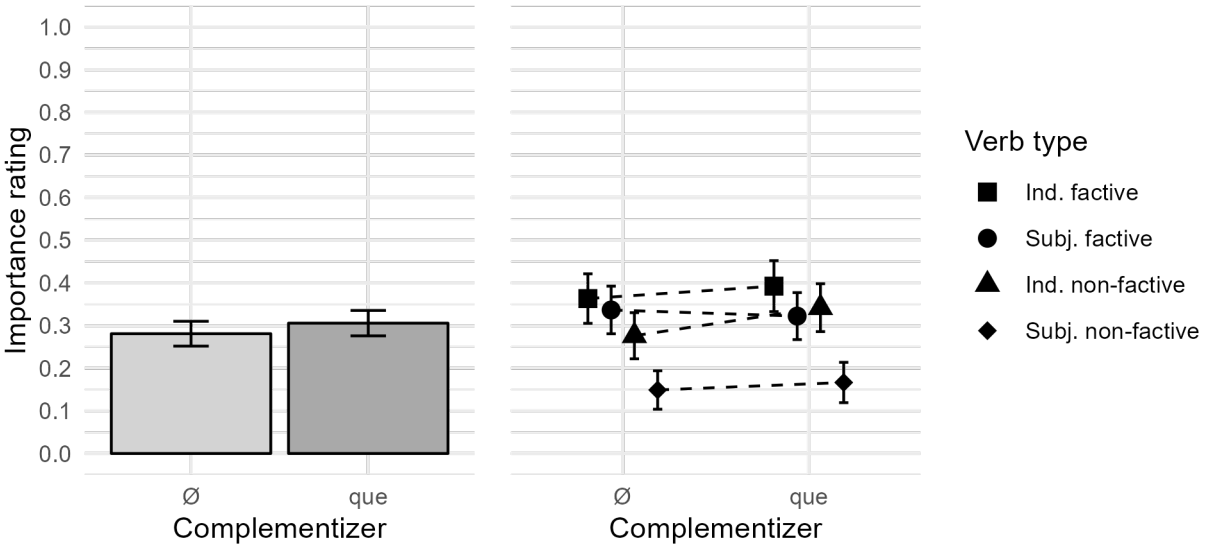


Figure 5.4: Importance ratings for at-issueness: Proportion of *sí* "yes"-responses and 95% confidence interval by experimental condition (Experiment 2). For all embedding verbs (left) and for the four verb types (right).

The importance ratings for at-issueness in Figure 5.4 are surprisingly low across the board. The proportion of *yes*-responses is below 0.5 overall, as well as for the four individual verb types. The indicative-taking non-factives should have received much higher ratings, given the assumption that these verbs may embed at-issue content (see Chapter 3.1.1 for details on this assumption). One possible explanation for the across the board low ratings is the fact that the experimental items featured coordinate embedded

clauses. This means that in an experimental item like (17), repeated from (9) and (15) above, there are two coordinate CCs of equal value: one, *el rey fuma cada día* 'the king smokes every day', and two, *la reina fuma a veces* 'the queen smokes sometimes'. The formulation of the at-issuness diagnostic to this item in (18) carries a uniqueness inference triggered by the superlative *lo más importante* 'the most important thing'. This inference may have made participants look for one single "most important thing" in the items, which may have clashed with the presence of two syntactically and semantically equal conjuncts of overall equal importance. It would therefore not be appropriate to single out one conjunct, i.e. the elliptical one in the experiment, as the sole "most important" one. I return to this possible explanation of the low importance ratings in Section 5.2.2, where I compare fillers with and without clausal coordination in this regard.

(17) **Sample item, Experiment 2** (same as in example 9 and 15 above)
María: Santiago cree que el rey fuma cada día y {**que** / Ø} la reina, a veces.
María Santiago thinks that the king smokes every day and that the queen at times
'María: Santiago thinks that the king smokes every day and that the queen smokes sometimes.'

(18) **Importance diagnostic (for at-issueness)**
Lo más importante que dice María es que la reina fuma a veces.
the most important that says María is that the queen smokes-IND at times
'The most important thing that María is saying is that the queen smokes sometimes.'

Finally, the reportiveness ratings are in Figure 5.5. Once more, the indicative-taking non-factives should receive a relatively high proportion of *yes*-responses, given that this type contains verbs of saying and thinking. While these verbs do receive the highest *yes*-proportion, the two factive types receive similarly high *yes*-proportions, which do not seem to suggest a categorical difference between these three verb types. Only the subjunctive-taking non-factives receive the expected low proportion of *yes*-responses. This result is surprising because it seems to suggest that emotive verbs (i.e. the subjunctive-taking factives) and verbs of knowledge and perception (i.e. the indicative-taking factives) may quite easily receive reportive ratings. Alternatively, the reportiveness diagnostic may have not been sensitive enough to be indicative of actual reportiveness. The fact that the subjunctive-taking non-factives received relatively low reportiveness ratings may have to do with the negative inference they might carry regarding their CC, as discussed above.

In sum, the results of Experiment 2 do not straightforwardly provide any new, positive information on the effect of *que* in gapping on interpretation. All three diagnostics deliver null results across the board. Neither the pooled data for all verbs, nor the data of the four individual verb types display any noticeable effect of *que* in gapping. Under the assumption that the diagnostics accurately measured those semantic/pragmatic properties that they were intended to measure, i.e. assertion, at-issueness, and reportiveness, these results mean that *que* does not carry meaning that causes an interpretation along the lines of these properties.

The assumption that the diagnostic delivered accurate measurements does not seem to be accurate in light of the results: the results of the four verb types (independently of *que*) show that none of the diagnostics has returned results that were entirely according to expectations. For this reason it is likely that the diagnostics, on the whole, inadequately measured the three properties in question. The following

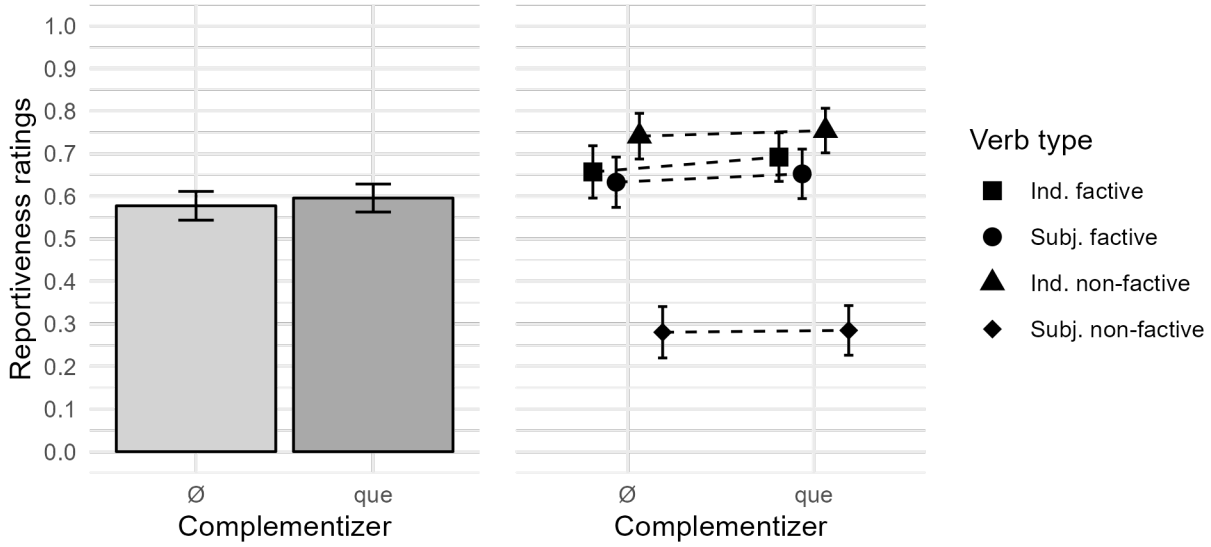


Figure 5.5: Reportiveness ratings: Proportion of *sí* "yes"-responses and 95% confidence interval by experimental condition (Experiment 2). For all embedding verbs (left) and for the four verb types (right).

section checks the diagnostics once more, this time on the basis of the filler items of Experiment 2. The results of this check will confirm the observation in this section that at least some of the diagnostics were likely inadequate.

5.2.2 A critical examination of the diagnostics

As mentioned in the introduction to Section 5.2, each list of Experiment 2 contained 16 filler items in addition to the 24 experimental items. These filler items were meant to fulfil a number of different functions at once. For this reason, some individual filler items featured distinct different structural properties at once, which aimed at checking different aspects of the diagnostics. In this regard, the design of the filler items represents an extension of the aforementioned overall "exploratory spirit" of Experiment 2. One straightforward function of the filler items was to distract participants from the structure that was the main focus of Experiment 2, i.e. embedded gapping with and without *que*. In addition to this distracting function, the filler items aimed to investigate whether the three diagnostics for assertion, at-issueness, and reportiveness were working correctly. To this end, there were four distinct types of filler items, which differed in the structural building blocks they featured. There were four individual filler items in each of the four types, which makes the total of 16 filler items.

One type of filler items featured a phrase headed by the preposition *según* 'according to'. This phrase fulfills a reportive function without clausal embedding, see (19) for a sample filler item of this type. This type of filler item was thus primarily intended to check the reportiveness diagnostic: the reportive meaning of the *according to* PP should result in a high proportion of *yes*-responses for the reportiveness diagnostic, which targeted the part of the filler item that did not include the *according to* PP. See (20) for the statement that participants responded to in the reportiveness diagnostic for the item in (19).

(19) **Sample of *according to*-control item (with PP headed by *según* 'according to')**

María: Según el metre, la cocinera prepara la sopa y los camareros ponen las
María according.to the head.waiter the cook prepares the soup and the waiters set the
mesas.
tables
'María: According to the head waiter, the cook prepares the soup and the waiters set the tables.'

(20) **Sample of *according to*-control item (with PP headed by *según* 'according to')**

El metre le contó a María que la cocinera preparaba la sopa.
The head.waiter her told to María that the cook prepared the soup
'The head waiter told María that the cook prepares the soup'

The second type of filler items involved sentences with nominal appositions, such as *un académico de gran talento* 'a talented scholar' in (21). In these sentences, the diagnostics for assertion and at-issueness targeted the content of the apposition, see (22) and (23), respectively. The content of an apposition is usually not asserted (see the discussion on apposition in the introduction to this section above). Consequently, the *possible that not* diagnostic should receive a low proportion of *yes* ratings (under the assumption that this diagnostic indeed measures assertion, which it might not; see Section 5.2.1). Additionally, the apposition fillers should receive a low *yes*-proportion for the importance diagnostic, because the content of an apposition is usually not at issue.

(21) **Sample of apposition filler item**

María: Luis, un académico de gran talento, enseña gramática a Samuel y a David, latín.
María Luis a academic of great talent teaches grammar to Samuel and to David Latin
'María: Luis, a talented scholar, teaches grammar to Samuel and Latin to David.'

(22) ***Possible that not*-diagnostic**

Es posible que Luis no sea un académico de gran talento.
is possible that Luis NEG is.SUBJ a academic of great talent.
'It is possible that Luis is not a scholar of great talent.'

(23) **Importance diagnostic**

Lo más importante que dice María es que Luis es un académico de gran talento.
the most important that says María is that Luis is a academic of great talent.
'The most important thing that María is saying is that Luis is a scholar of great talent.'

The third type of filler items involved sentences with the phrase *cómo ya sabéis* 'as you$_{PL}$ already know' as in (24). This phrase was intended to mark the content of the sentence as discourse-old. These filler items were supposed to check the diagnostic for assertion, since assertion is a means to introduce new information to discourse (see Chapter 3.1.1). The proportion of *yes*-answers to the *possible that not*-diagnostic is expected to be low for these filler items. However, due to an oversight in experimental design, the measurement that these filler items provide for the *possible that not*-diagnostic are likely not informative: three of the four filler items of this type involved clausal embedding (i.e. the filler item in 24 is the only one that does not involve clausal embedding), and the diagnostic targeted the content of the embedded clause, rather than that of the main clause. The responses that participants gave for the *possible that not*-diagnostic for these verbs are thus more likely to be a function of the embedding predicates in these sentences, rather than a function of the *as you know* phrase.

(24) **Sample of *as you know*-filler item (with phrase *como ya sabéis* 'as you already know')**
María: Como ya sabéis, el albañil llegó tarde al trabajo y el fontanero, bebido.
María as already know-2PL the bricklayer arrived late to work and the plumber drunk
'María: As you already know, the bricklayer arrived late at work and the plumber drunk.'

Finally, the fourth type of filler items were 'plain' sentences, i.e. sentences with no special structural or semantic feature as the other filler items. One of these 'plain' filler items is in (25). For these items, the assertion and at-issueness diagnostics targeted the content of the whole sentence. The proportions of *yes*-responses should thus be high for the importance diagnostic, since the content of the clause is very much at issue, and high for the *possible that not*-diagnostic, since María could be lying or mistaken about the content of her utterance.[9]

(25) **Sample of "plain" filler item (without any special structural or semantic feature)**
María: Inés restaura la antigua pintura de la iglesia local sin ningún tipo de ayuda
María Inés restores the old painting of the church local without any type of help
profesional.
professional
'María: Inés restores the old painting in the local church without any professional help.'

As you can see in the examples above, some of the filler items contained clausal coordination (examples 19 to 24), a subset of which featured gapping (examples 21 and 24). Of the 16 filler items, 13 contained clausal coordination, three of which were elliptical. I will return to the difference between coordinate and simple control items below. Ellipsis will play no role in the discussion of the filler items since ellipsis is unlikely to play a role for the ratings. Furthermore, five of the filler items contained clausal embedding. Three of these filler items with embedding were of the *as you know*-type (see the discussion above). The other two filler items with embedding were "plain" fillers. For the fillers with embedding, the diagnostics targeted the content of the embedded clause rather than that of the matrix clause, which leads to the problems I discuss above for the *as you know*-fillers: the resultant ratings are likely dependent on the embedding verb rather than of the whole filler sentence as such. This problem limits the degree to which the ratings for the "plain" fillers are informative.

In the questionnaire, each filler item was followed by three statements, which participants responded to by means of binary *yes/no* judgments. This procedure was identical to that of the experimental items, see the introduction to Section 5.2. For the fillers with *according to*, the three statements contained the same three diagnostics as those of the experimental items, i.e. the *possible that not* diagnostic, the importance diagnostic, and the reportiveness diagnostic. Thus, for the filler item in (19) above (reproduced in 26 below), participants responded to the three statements in (27) to (29). For the filler items containing clausal coordination, the statements targeted the first conjunct (i.e. *la cocinera prepara la sopa* 'the cook prepares the soup' in 19 above). The reason for this aspect of the design was to balance out the experimental items that targeted the elliptical, i.e. the second, conjunct. However, it may have lead to inaccurate results in for the importance diagnostic, since in a structure with coordinate clauses, both clauses are equally important.

[9]As will be seen by the surprisingly low proportion of *yes*-responses for the *possible that not*-diagnostic (see below), participants on the whole did not seem to consider the possibility that María was lying or mistaken about the content of her utterance.

I return to this point below, see also the similar remarks on the interaction of clausal coordination with the importance diagnostic in the discussion of the experimental items in Section 5.2.1.

(26) **Sample of *according to*-control item** (same as 19 above)
María: Según el metre, la cocinera prepara la sopa y los camareros ponen las
María according.to the head.waiter the cook prepares the soup and the waiters set the
mesas.
tables
'María: According to the head waiter, the cook prepares the soup and the waiters set the tables.'

(27) ***Possible that not* diagnostic**
Es posible que la cocinera no prepare la sopa.
is possible that the cook NEG prepares the soup
'It is possible that the cook does not prepare the soup.'

(28) **Importance diagnostic**
Lo más importante que dice María es que la cocinera prepara la sopa.
the most important that says María is that the cook prepares the soup
'The most important thing that María is saying is that the cook prepares the soup.'

(29) **Reportiveness diagnostic**
El metre le contó a María que la cocinera preparaba la sopa.
The head.waiter her told DOM María that the cook prepared the soup
'The head waiter told María that the cook prepared the soup.'

The other three types of filler items also tested statements with the *possible that not*-diagnostic and the importance diagnostic. Since reportiveness was only relevant to the *according to*-controls, the other fillers did not come with a statement targeting reportiveness like (29) above, but instead came with a distractor statement. The distractor statement was a repetition of the direct speech in the filler item, but with the speaker (*María*) as the subject instead of the subject in María's utterance. For an illustration, see the distractor statement in (31) to the "plain" filler item in (30), where the original subject *Inés* was replaced by the speaker *María*. All statements of these distractor type were false given the speaker's utterance (none of María's statements involved herself), which predicted that participants consistently indicated *no* as their response.[10]

(30) **Sample of "plain" control item** (same as example 25 above)
María: Inés restaura la antigua pintura de la iglesia local sin ningún tipo de ayuda
María Inés restores the old painting of the church local without any type of help
profesional.
professional
'María: Inés restores the old painting in the local church without any professional help.'

(31) **Distractor statement**
María restaura la antigua pintura de la iglesia local sin ningún tipo de ayuda
María restores the old painting of the church local without any type of help
profesional.
professional
'María restores the old painting in the local church without any professional help.'

[10] Four of the 80 participants that completed the questionnaire had answered more than two of the twelve distractor statements with *sí* 'yes' (i.e. incorrectly). The data of these four participants did not enter the analysis of Experiment 2 (i.e. neither the data in this section nor the data from the experimental items in Section 5.2.1 features data from these four participants).

The proportion of *yes*-responses for the three statements to each filler item, illustrated for each of the four types of filler items, is in Figure 5.6. This figure also provides an overview over which diagnostics applied to which filler types. As expected, the *yes*-proportion for the distractors is very low for the three filler types that came with distractor statements. Also as expected, the reportiveness diagnostics yielded a high *yes*-proportion for the *according to*-fillers. This high proportion indicates that a phrase headed by *según* ´according to' indeed receives a reportive interpretation. In this regard, the diagnostic for reportiveness seems to have worked as expected. However, in hindsight it would probably have been good to check whether the diagnostic accurately yields low *yes*-proportions for non-reportive sentences. Recall that emotive verbs and knowledge/perception verbs (i.e. subjunctive-taking and indicative-taking factives) received surprisingly high *yes*-proportions in the reportiveness diagnostic in the experimental results (Section 5.2.1).

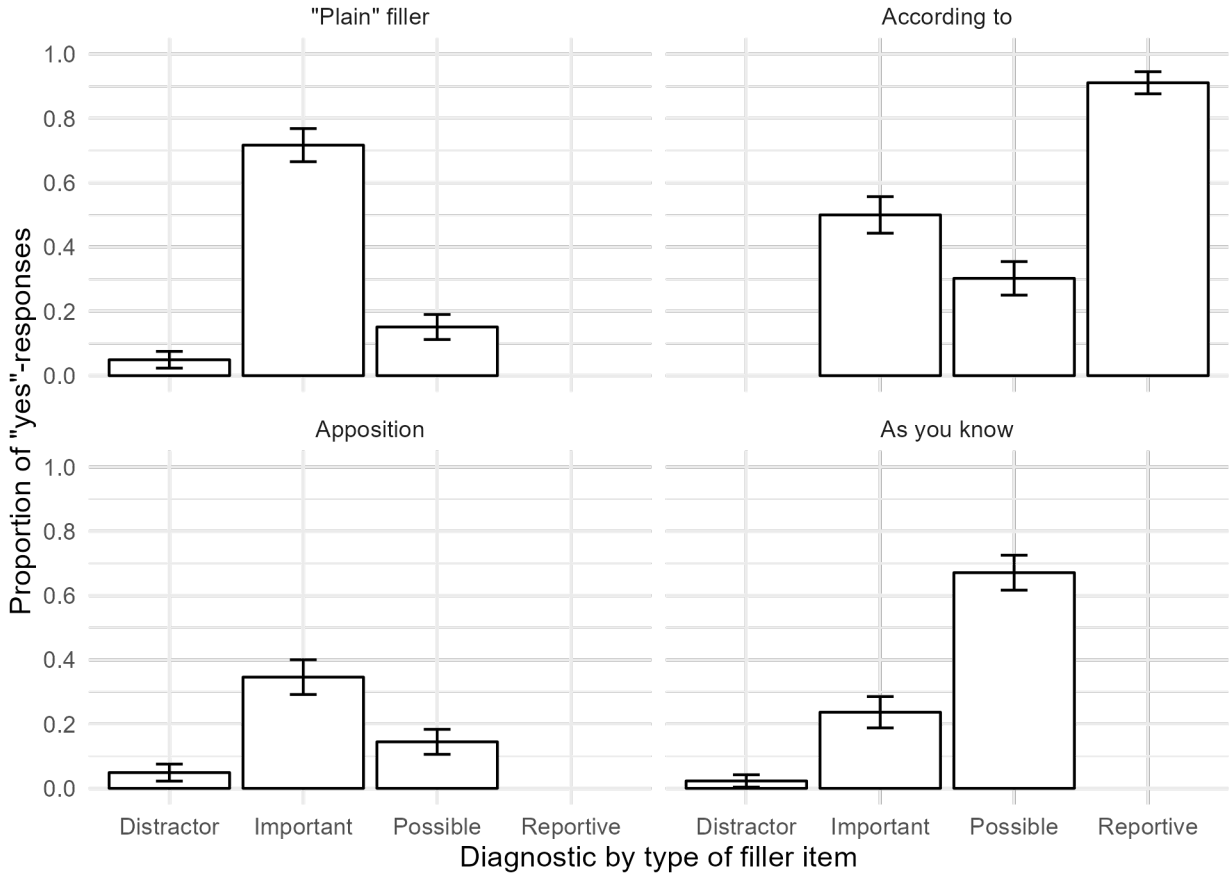


Figure 5.6: Proportion of *yes*-responses with 95% confidence interval by diagnostic and type of filler item (Experiment 2).

The other two diagnostics, importance and *possible that not*, show more differential patterns in the four filler types. Due to the oversight discussed above, due to which the diagnostics in the "plain" and *as you know*-filler items targeted the content of embedded clauses, these two filler types are not informative with respect to the importance and *possible that not*-diagnostics. For the apposition fillers, both the importance and the *possible that not*-diagnostics yielded low proportions of *yes*-responses. This indicates that the diagnostics accurately picked up on the fact that the content of the appositions (which were the targets

for the diagnostics) were not at issue (hence the low *yes*-proportion for importance) and not asserted (hence the low *yes*-proportion for assertion).

The importance diagnostic for the *according to*-fillers indicates a roughly equal proportion of *yes* and *no* responses. This might indicate that in these fillers, the interpretation of what the "most important thing" is is variable between the content of the sentence (indicated by *yes*-responses) or the source of that content encoded in the *according to* PP (indicated by *no*-responses). More interestingly, the *according to*-fillers also show relatively low *yes*-proportion for the *possible that not*-diagnostic (albeit larger than for the apposition fillers), which given the reportive nature of these fillers is unexpected. These fillers do not contain any elements that would indicate non-assertion of the content that the diagnostics target, i.e. the whole sentence without the *according to* PP. This result casts doubt on the assumption that the *possible that not*-diagnostic accurately measured assertion.

This doubt concerning the *possible that not*-diagnostic is augmented by the fact that the fillers contain no reliable data for the *possible that not*-diagnostic applied to simple declarative clauses, i.e. clauses encoding simple assertions. Such clauses should receive a high proportion of *yes*-ratings. Without checking the *possible that not*-diagnostic in this direction, we cannot say for sure whether it is accurate. Additionally, this diagnostic suffers from a number of conceptual problems anyway (see above in this section), which the data from the fillers similarly does not resolve. We should therefore remain (very) sceptical about whether this diagnostic has accurately measured assertion or not.

As a last point, recall from the results for the experimental items (Section 5.2.1) that the importance diagnostic (*Is the most important thing that María is saying the CC?*) returned surprisingly low proportion of *yes*-responses for all four verb types. These results go against expectations as the indicative-taking non-factives (verbs of saying and thinking) should, as per the assumptions in Chapter 3.1.1), be able to embed at issue content, and therefore show a higher proportion of *yes*-responses. In the discussion of these results, I speculated that the low *yes*-proportion may be due to the fact that the diagnostic targets one of two coordinate clauses, which independently of the type of embedding verb are of equal importance. Consequently, the content of one conjunct cannot be the "most" important one.

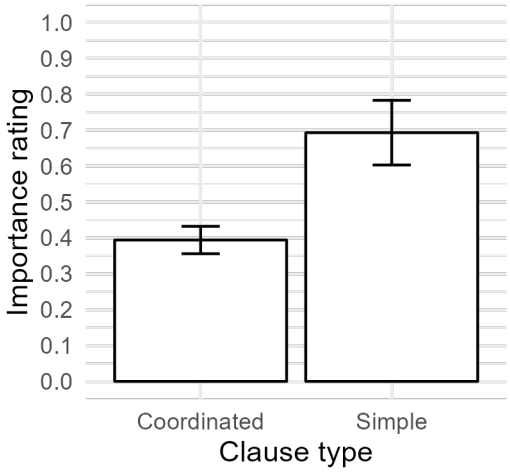


Figure 5.7: Proportion of *yes*-responses with 95% confidence interval for simple vs. coordinated filler items (Experiment 2). Collapsed data from all filler types.

The filler items contained both simple and coordinate clauses. They thus provide an appropriate testing ground for my speculation: filler items with coordinate clauses should receive a low *yes*-proportion in their importance ratings, while simple clauses should receive a higher *yes*-proportion in their importance ratings (since the content of that simple sentence *is* the most important part of the speaker's utterance). Figure 5.7 shows the *yes*-proportions of the simple and coordinated filler items. It shows that my speculation is borne out: filler items with coordinate clauses show a much smaller *yes*-proportion than filler items without clausal coordination. The low proportion of *yes*-responses in the experimental items (Section 5.2.1) is thus likely not indicative of the real underlying interpretation of at-issueness for the four verb types, but instead an artifact of the clausal coordination in the experimental items.

5.3 Conclusion

The two experiments in this chapter approached the meaning of the complementizer *que* 'that' in Spanish globally embedded gapping. The experimental results strongly suggest that the hypothesis that *que* has a (strong) effect on interpretation is inaccurate. In Experiment 1, I applied the *certain that*-diagnostic for (factive) presupposition to gapping with vs. without *que*. The presence of *que* makes no difference in the interpretation of certainty. In Experiment 2, I applied three innovative diagnostics to the same materials that targeted the semantic/pragmatic properties assertion, at-issueness and reportiveness. Once again, *que* showed no discernible effect on either of these properties.

While a critical examination of the diagnostics in both experiments show that the diagnostics most likely did not accurately measure the properties they were intended to measure (i.e. factivity in Experiment 1, assertion, at-issueness and reportiveness in Experiment 2), I maintain that the results of these two experiments strongly suggest that *que* in gapping has no influence on interpretation along the general lines of these four properties. Despite the considerable inaccuracies in the diagnostics, the measurements still generally relate to the crucial aspects of meaning that I set out to investigate. Perhaps the diagnostics do so in an indirect and non-ideal way, but they still indicate notable differences between the verb types (Section 5.2.1) and, partially, in the filler items (Section 5.2.2). These diagnostics thus each seemed to be sensitive to *some* aspect of meaning, however closely related these aspects may have been to the intended aspects. For this reason, I would still confidently expect at least one of the four verb types to show a *que*-effect in at least one of the four total diagnostics applied in the two experiments in this chapter, had the complementizer had the strong effect on interpretation that I hypothesized in Chapter 3. The fact that the experimental results of this chapter are entirely null results thus strongly indicates that the hypothesis that *que* in gapping influences interpretation is wrong.

Unfortunately, I see no possibility to give the null results and the negative conclusions of this chapter a positive spin. The sensitivity of *que* in gapping to the type of embedding verb that I observed in Chapter 3 must remain unaccounted for, at least for now. However, I believe that the content of this chapter is still very much relevant to the study of *que* in gapping, since it provides quantitative data on my hypothesis on the meaning of *que*. This data has in turn served as an empirical basis for testing, and ultimately rejecting, this hypothesis – as is expected in rigorous empirical inquiry. This constructive view of the contribution of this chapter is the positive aspect with which I would like to conclude this chapter, and indeed the bulk of my dissertation. The next chapter contains a general summary and conclusion of all chapters.

6 Summary and conclusion

This dissertation has investigated verb ellipsis in coordinate clauses (gapping and, to a lesser extent, stripping) in the context of clausal embedding. At the center of investigation have been two structural configurations that involve verb ellipsis and clause-embedding, which I termed "globally embedded" and "locally embedded" verb ellipsis, see (1) and (2) for gapping, respectively. In Chapter 2, I compared these structures with respect to their ability to host a complementizer. In English, the complementizer *that* is ungrammatical in both globally and locally embedded gapping, as shown in (1) and (2).

(1) **Globally embedded gapping**
John thinks that Mary practices the piano and (***that**) Sarah ___ the clarinet.

(2) **Locally embedded gapping**
Mary practices the piano and John thinks (***that**) Sarah ___ the clarinet.

This ungrammaticality of the complementizer in both structural configurations also shows in other languages (Dutch, German, French), but is by no means universal. In Chapter 2.3, I list a number of typologically diverse languages that allow complementizers in embedded gapping. In the investigation into embedded gapping in these languages, a pattern seemed to emerge: if a language allows complementizers in locally embedded gapping, it also allows complementizers in globally embedded gapping (and vice versa). I captured this observation by proposing a generalization in form of the Typological Conjecture in (3).

(3) **Typological Conjecture on embedded gapping and stripping**
Any constraint on a complementizer specific to ellipsis that operates in local embedding also operates in global embedding (and vice versa).

I was (and still am) hesitant to present this generalization as more than a conjecture (like a contraint or universal), since cross-linguistic data on locally and globally embedded verb ellipsis is sparse and partially unreliable. I believe, however, that the Typological Conjecture may serve as a starting point into the cross-linguistic investigation of embedded ellipsis. In Chapter 2.4, I show how such an investigation might proceed, using the specific example of Russian. In that chapter, I investigate locally embedded gapping and stripping with the polarity particle *tože* 'too'. For the grammaticality status of gapping specifically, there is unclear and contradictory introspective data in the theoretic literature. I clear up this difficult empirical situation by providing quantitative judgment data that shows that gapping is degraded, but not outright ungrammatical in local embedding.

More to the point from the perspective of the Typological Conjecture, I compare these data on locally embedded gapping and stripping to data on globally embedded gapping and stripping from Bonke

(forthcoming). The comparison shows that both these structures are subject to the same constraints on discourse relations: if the conjuncts are in a contrastive relation, ratings indicate a degradation in acceptability. If the conjuncts are in a parallel, non-contrastive relation, there is no such degradation.

The remainder of the dissertation was concerned with Spanish. In this language, the empirical situation on locally embedded verb ellipsis with the complementizer *que* 'that' is comparatively good. There are two quantitative contributions, a corpus study on stripping (Garcia-Marchena 2018), and an acceptability study on gapping (Bîlbîie & de la Fuente 2019). In Chapter 2.5, I provide an in-depth summary (and critique) of Bîlbîie & de la Fuente's (2019) acceptability study. The main result of that acceptability study is that the acceptability of locally embedded gapping is subject to constraints on the embedding verb, which Bîlbîie & de la Fuente (2019) analyse in terms of factivity. With non-factive embedding verbs (i.e. verbs of saying and thinking), locally embedded gapping is near fully acceptable. With factive embedding verbs, ratings indicate differential degrees of degradation in acceptability, depending on the specific subtype of embedding verb. With semi-factives (i.e. knowledge and perception verbs), the degradation is moderate, with true factives (i.e. emotives), the degradation is more pronounced. For none of the types, however, is the degradation so large as to indicate outright unacceptability.

The Typological Conjecture now predicts that we should find the same verb type constraints on *que* in globally embedded gapping. I test this prediction by means of the acceptability study in Chapter 3. In a series of seven acceptability judgment experiments, I find that *que* in globally embedded gapping is generally acceptable – a finding which goes against impressionistic judgments in the literature that claim the opposite (Fernández-Sánchez 2021). I relate these findings to the theoretic issue of conjunct size in gapping: the fact that a gapping conjunct may contain the head of a clause (i.e. the complementizer) provides evidence that gapping, at least in Spanish, must involve the coordination of clause-sized constituents (e.g. CPs).

However, the results of Chapter 3 do not indicate that *que* in gapping is fully acceptable across the board. Instead, *que* is subject to the same constraints on the embedding verb as locally embedded gapping: Verbs of saying and thinking (i.e. Bîlbîie & de la Fuente's 2019 non-factives) show hardly any degradation with *que*, while the two factive sub-types show varying degrees of degradation with *que*. As in Bîlbîie & de la Fuente's (2019) experiment, knowledge and perception verbs (i.e. semi-factives) show a smaller degree of degradation than emotive verbs (true factives). The degradation was not so substantial for any of the types of verbs to indicate outright ungrammaticality. Thus, the acceptability patterns for these three verb types are identical in locally and globally embedded gapping, as predicted by the Typological Conjecture. However, my acceptability study contained a further verb type that Bîlbîie & de la Fuente (2019) did not test. This verb type consisted of non-factive verbs that were not verbs of saying and thinking, i.e. desideratives and verbs with implicit negation. In terms of the acceptability of *que*, this verb type patterned with the two factive verb types in showing a similar kind of degradation in acceptability.

These results suggested that while the same constraint may be at play in both locally and globally embedded gapping, factivity is unlikely to be the crucial property in that constraint. To capture the acceptability patterns of *que* in gapping, I sketched a structural account based on an articulated left periphery and the structural truncation of (some) complement clauses (Chapter 3.5). One building block of this analysis was the assumption that *que* in gapping is indicative of a structure that carries its own meaning. This meaning, as per assumption, harmonizes with the meaning of the verbs of saying and thinking, which

therefore show no degradation with *que*. The meaning of the other verb types, on the other hand, clash with the meaning indicated by *que*, which may be repaired by shifting the interpretation of the verbs towards the meaning indicated by *que*, at the cost of some acceptability. In this way, we may arrive at the gradual degradation caused by *que* across embedding verbs and verb types.

The obvious question that arose from this reasoning was what meaning *que* might be indicative of. The discussion in Chapter 3.5 raised four potential candidates: the aforementioned factivity, assertion, at-issueness, and reportiveness. The reason to keep factivity under consideration despite the fact that the results of Chapter 3 suggested that factivity played no central role in the acceptability of *que* stemmed from the way I had categorized the embedding verbs into factive and non-factive. The decision as to which verbs counted as factive and which verb counted as non-factive proceeded on my personal impressionistic judgments, which led to a considerable empirical problem for the conclusions in Chapter 3. Given my limited practical knowledge of Spanish, this procedure may have led to inaccuracies in the categorization of verbs. To amend this problem, I conducted a rating study in Chapter 4, which featured a large sub-set of the embedding verbs in the acceptability study. In this study, I collected certainty ratings for the verbs using the *certain that*-diagnostic in polar questions (see Degen & Tonhauser 2022 among others). These ratings are indicative of the degree to which the embedding verbs are interpreted as factive. The design of the study was modelled after Degen & Tonhauser's (2022) study on English.

The results of this study revealed a number of interesting properties of Spanish clause-embedding verbs, independently of the question of *que* in gapping. This study thus represents a standalone contribution to the semantics and pragmatics of clause embedding. One of these interesting properties, and the main result of this study, is that the certainty patterns for the individual verbs parallel Degen & Tonhauser's (2022) findings on English in that there is no abrupt difference in certainty ratings between (presumed) factive and non-factive verbs, but rather a gradual incline that suggests no distinct categories of embedding verbs as far as factivity is concerned (such as the overall distinction of factive vs. non-factive, or semi-factive vs. true factive as sub-types within the factives). Further interesting results were the fact that mood (i.e. indicative vs. subjunctive in the embedded clause) does not seem to influence factivity as strongly as sometimes assumed (Chapter 4.3.1 explicitly discusses Rivero 1971 in this regard), and that verbs with implicit negation received unexpectedly high certainty ratings, even though I had classed them as non-factive. Within the gradual cline of factivity, however, my categorization of the verbs into non-factive and factive in Chapter 3 had been relatively accurate: most of the verbs I had classed as non-factive tended towards the lower end of the certainty scale, and most of the verbs I had classed as factive tended towards the higher end of the scale.

Coming back to the question of *que* in gapping, the certainty ratings recorded in Chapter 4 provided a new way of investigating whether it was factivity that determined the degradation of *que*. The comparison of the experimental data from Chapters 3 (acceptability) and 4 (factivity) revealed that on the basis of individual verbs, there is a positive, statistically significant correlation of factivity and *que*-degradation, leading some credence to the hypothesis that factivity determines the acceptability of *que* in gapping after all. However, the correlation was only weak, and subsequent investigation revealed that the correlation was driven by the verbs of saying and thinking. Without the verbs of this type, the correlation disappears. This observation strengthened the conclusion of Chapter 3, which was that the verbs of saying and thinking specifically (rather than all non-factive verbs) easily allow for *que* in gapping, in contrast to all other

verb types, including the two types of factives and the other non-factives (desideratives and verbs with implicit negation).

In the last chapter, Chapter 5, I tried to capture the hypothesized effect of *que* in gapping directly, i.e. by applying diagnostics to globally embedded gapping with *que* vs. without *que*. Chapter 5 features two rating experiments aimed to explore whether *que* affects the interpretation of factivity, assertion, at-issueness, and reportiveness. One experiment applied the *certain that*-diagnostic to the gapping materials from Chapter 3, to see whether *que* would lead to a decrease in certainty. Such an effect may have been interpreted as evidence for a non-factive meaning indicated by *que* in gapping. However, in the results *que* showed no meaningful effect on the interpretation of certainty, which indicated once more that *que* unlikely interacts with factivity in gapping. The second experiment in Chapter 5 used innovative diagnostics to explore the interaction of *que* and the interpretation of assertion, at-issueness and/or reportiveness. An attempt at validating the diagnostics based on the results for the individual verb types and control items revealed that the accuracy of these diagnostics was questionable. This means that they likely did not measure assertion, at-issueness or reportiveness. However, since the diagnostic did seem to pick up on some aspects of meaning (i.e. there were noticeable differences in the ratings for different verb types and within the respective control items), I still chose to include this experiment in the dissertation. The results of this experiment, as far as *que* in gapping is concerned, indicated no noticeable effect of *que* in general, or within any of the investigated verb types. These results led me to conclude that my initial hypothesis, i.e. that *que* is indicative of a meaningful structure, is likely incorrect.

All in all, this dissertation has proposed a generalization, in form of the Typological Conjecture, on the relationship between two configurations of embedded gapping: globally and locally embedded gapping. Using the Typological Conjecture as a starting point, the dissertation has provided a wealth of data on three different languages (Appendix A features two experiments on German globally embedded gapping), which either serves to illuminate a difficult empirical situation (i.e. the Russian data on locally embedded gapping in Chapter 2.4), or outright disproves incorrect impressionistic judgments in the literature (i.e. the Spanish data of Chapter 3). It has provided further novel data on the interpretation on Spanish clause-embedding verbs in terms of factivity (Chapter 4). The data in that chapter provides numerous angles of interpretation, from the way verbs ought to be categorized into types, via the pragmatic properties of mood, to cross-linguistic investigations of factivity. Finally, the dissertation ended with an exercise of scientific rigour by attempting to provide quantitative evidence for hypothesized facets of meaning caused by syntactic structure (Chapter 5).

Appendix A: The complementizer *dass* 'that' in German embedded gapping

In this short chapter, I present the results of two acceptability judgment experiments on the German complementizer *dass* 'that' in globally embedded gapping, i.e. a structure like in (1). The results of these two experiments show that *dass* is unacceptable in this structure, which is in accordance with the introspective judgments in the literature (Hartmann 2000, Reich 2007, Repp 2009).

(1) **German**
Ich glaube, dass Peter mit seiner Frau nach Indien reist, und (***dass**) Martin mit seinen
I believe that Peter with his wife to India travels and that Martin with his
Kollegen in die Schweiz ___.
colleagues in the Switzerland
'I believe that Peter is travelling to India with his wife and that Martin is travelling to Switzerland with his colleagues.' (adapted from Hartmann 2000: 157 f., my gloss and translation)

The data presented in this chapter thus provides a contrast to Spanish, where the complementizer *que* 'that' is generally acceptable in globally embedded gapping (see Chapter 3 for details). In a similar way, the data constitutes a contrast to my study on Russian (Bonke forthcoming), where the complementizer *čto* 'that' shows degraded ratings, however without being outright unacceptable (see also Chapter 2.4) for some details). Against this background, it is worthwhile to include the present study on German because it illustrates the patterns of a language where the complementizer is genuinely unacceptable in globally embedded gapping.

Additionally, this chapter serves as a "sanity check" on the aforementioned introspective judgments from the theoretic literature that claim that the complementizer is unacceptable in German. This check is relevant since as far as introspective judgments in the theoretical literature are concerned, Spanish, Russian and German should pattern alike. For all three of these languages, the complementizer has been claimed to be ungrammatical, see the examples in (2) and (3) for Spanish and Russian, respectively.

(2) **Spanish**
Supongo que María traerá las bebidas y (***que**) Pedro ___ el postre.
suppose that María will.bring the drinks and that Pedro the dessert
'I suppose that Mary will bring the drinks and that Pedro will bring the dessert.' (Fernández-Sánchez 2021: 122 f.)

(3) **Russian**
Ja dumaju, čto Petr ezdit v Indiju a (***čto**) Pavel ___ v Švejcariju.
I think that Peter is.going to India and that Pavel to Switzerland
'I think that Peter is going to India and that Pavel is going to Switzerland.' (Repp 2009: 226)

For Spanish and Russian, these judgments are of course inaccurate, since gapping with the complementizer is at most degraded, but not outright unacceptable. Based on the inaccurate judgments in these two languages, one might suspect that there is a similar situation in German. Hypothetically, the ungrammatical judgment in the German example (1) might actually stem from a degradation of the complementizer, and not from outright unacceptability. As this chapter will show, this suspicion is unwarranted. The complementizer is indeed unacceptable in German globally embedded gapping.

In Section 1, I outline the design of the two experiments. Sections 2 and 3 contain the results of Experiments 1 and 2, respectively. Section 4 closes this chapter with a short discussion of the results and a conclusion.

1 Experimental design

The two experiments are identical in design to the Spanish Experiments 1 and 2 presented in Chapter 2. For details on the design, see that chapter. In this section, I will provide a description of the design and the experimental materials that covers all crucial aspects, but is less detailed than the description of the Spanish experiments. If you recall the design of the Spanish experiments (especially Experiments 1 and 1), you may want to skip this section altogether.

The two experiments had an identical factorial design. The first factor was the complementizer (present vs. absent *dass*), and the second factor was the type of ellipsis (gapping vs. stripping). The inclusion of stripping was an artifact of the design of the Spanish experiments after which the present experiments were modelled. There is no independent motivation for including this contrast between ellipsis types in German. In fact, *dass* should be equally unacceptable in gapping and stripping (see Chapter 2.1, which lays out the relevant facts for English and German). The two sample items in (4) and (5) illustrate all four resulting conditions of the target sentences in Experiment 1 and 2, respectively. The experimental items were largely translation equivalents to those of the Spanish experiments.[1]

(4) **Target sentence Experiment 1, all four conditions**
Thomas glaubt, dass der König jeden Tag raucht und {dass / Ø} die Königin {manchmal / auch}.
Thomas thinks that the king every day smokes and that the queen sometimes too
'Thomas thinks that the king smokes every day and that the queen smokes sometimes / that the queen smokes every day, too.'

(5) **Target sentence Experiment 2, all four conditions**
Sarah missbilligt, dass der König jeden Tag raucht und {dass / Ø} die Königin {manchmal / auch}.
Sarah disapproves that the king every day smokes and that the queen sometimes too
'Sarah disapproves of the fact that the king smokes every day and that the queen smokes sometimes / that the queen smokes every day, too.'

[1]One minor difference between the materials of the two experiments concerned the matrix subjects. To balance the ratio between male and female proper names, some of the matrix subjects were adjusted in the materials of Experiment 2. One of these changes is in (4) and (5), where the female male matrix subject of Experiment 2 *Sarah* substituted the male matrix subject *Thomas* of Experiment 1.

Experiment 1 tested embedding verbs of saying and thinking, which represent non-factive verbs. Experiment 2 tested emotive verbs, which represent factive verbs. Table 1 summarizes the embedding verbs.[2] Like with the distinction between gapping and stripping, the distinction between verb types had an independent motivation for the Spanish experiments. The same independent motivation does not apply in German. Since both verb types will turn out to show identical patterns w.r.t. *dass*, I will not go into detail on the differences between them. Instead, I will simply treat Experiment 2 as a replication/confirmation of Experiment 1.

Table 1: Embedding verbs in the two experiments.

Non-Factives (Experiment 1)		**Factives** (Experiment 2)	
German	*English*	*German*	*English*
behauptet	claims	bedauert	regrets
berichtet	tells	begrüßt	approves
denkt	thinks	bewundert	admires
deutet an	suggests	hasst	hates
erzählt	relates	kritisiert	criticizes
glaubt	thinks	lobt	praises
meint	considers	missbilligt	disapproves
mutmaßt	supposes	respektiert	respects
nimmt an	presumes	toleriert	tolerates
sagt	says	unterstützt	supports
schätzt	estimates	verabscheut	detests
vermutet	suspects	verteidigt	defends

The materials contained a set of 24 embedded clauses, which meant that each of the twelve verbs per experiment came with two of these embedded clauses. Unlike in the Spanish materials, all embedded verbs were in the indicative form, which is in accordance with the selectional preferences of the embedding verbs. Each of the resulting embedding verb-embedded clause combinations came in the four conditions in (4) and (5), for a total of 96 items. In each experiment, these 96 items were distributed across four lists, so that each participant judged every embedded clause once, and every embedding verb twice. Each list featured the same set of 36 control items (18 acceptable, 18 unacceptable), for a total of 60 items per list. Participants judged each item using a slider on a horizontal scale, whose endpoint were labelled *total unnatürlich* 'totally unnatural' and *total natürlich* 'totally natural'. The output was coded with integers from 1-100.

Each item was preceded by one to two sentences of context, which prepared the lexical content of the item, and followed by a *yes-no* verification question on the following page in the questionnaire. The verification question was there to ensure that the participants read the items carefully and got the intended interpretation. This form of attention check was particularly important for the items without *dass*, in which the subject of the elliptical conjunct is structurally ambiguous between the desired embedded interpretation (e.g. *the queen smokes every day, too* in the stripping variants of 4/5), or the non-desired matrix

[2]The English translations of the verbs in Table 1 indicate the translation equivalence with the verbs of the Spanish experiments, see Chapter 3, Table 3.1. For this reason, the translations of some of the translations in Table 1 provide only a rough equivalence with English (especially in the non-factive column).

interpretation (*the queen also thinks that the king smokes every day*). The context and overall design of the experiments further aided the participants in obtaining the intended interpretation by strengthening the semantic parallelism between antecedent and remnant subject (i.e. king and queen). Items for which the verification questions received an incorrect answer did not enter the analysis.

The questionnaires were hosted online on SoSciSurvey.de (Leiner 2024). For the statistical analysis, I used the packages `afex` (Singmann et al. 2022) and `lmer` (Kuznetsova et al. 2017) for `R` (R Core Team 2023) to fit linear mixed effects models. The results tables were computed with the function `summarySEwithtin()` from the package `Rmisc` (Hope 2022), which removes inter-subject variability. I used `ggplot2` (Wickham 2016) for the figures.

2 Experiment 1: Non-factives

Experiment 1 tested gapping and stripping under verbs of saying and thinking (i.e. non-factives). In (6), there is a sample item with the embedding verb *glaubt* 'thinks' in all four conditions (same as the sentence in 4 above).

(6) **Target sentence Experiment 1**
Thomas glaubt, dass der König jeden Tag raucht und {dass / Ø} die Königin {manchmal / auch}.
Thomas thinks that the king every day smokes and that the queen sometimes too
'Thomas thinks that the king smokes every day and that the queen smokes sometimes / that the queen smokes every day, too.'

35 participants completed the questionnaire, all of whom were from Prolific.com. The data from 8 bilingual participants did not enter the analysis, leaving 27 participants in the data set. The results are in Table 2 and Figure 1. The conditions with *dass* receive very low ratings, which are on a level with the unacceptable control items. The conditions without *dass* receive high ratings, approaching the mean rating for acceptable control items. Within the condition without *dass*, stripping received slightly lower ratings than gapping.

Table 2: Experiment 1 (gapping vs. stripping under non-factives): Mean ratings per experimental condition and control items (*N* = number of observations, *SD* = standard deviation, *SE* = standard error, *CI* = 95% confidence interval)

Ellipsis Type	Complementizer	N	Mean Rating	SD	SE	CI
Gapping	*dass*	154	19.09	21.17	1.71	3.37
Gapping	Ø	155	83.82	19.68	1.58	3.12
Stripping	*dass*	155	16.20	17.25	1.39	2.74
Stripping	Ø	159	77.84	22.94	1.82	3.59
High acceptability controls		381	88.39	28.70	1.47	2.89
Low acceptability controls		379	17.29	33.00	1.69	3.33

I think the comparison of the experimental conditions to the acceptable and unacceptable controls provides a sufficiently informative picture on the acceptability of *dass* in gapping. For sake of completeness, I also report the results of a linear mixed model with the data from the experimental items (i.e. without the controls). The model contained the two factors ellipsis type and complementizer, as well as their interaction. As expected, the effect of the complementizer was significant ($b = 31.69, SE = 1.96, t = 16.16, p <$

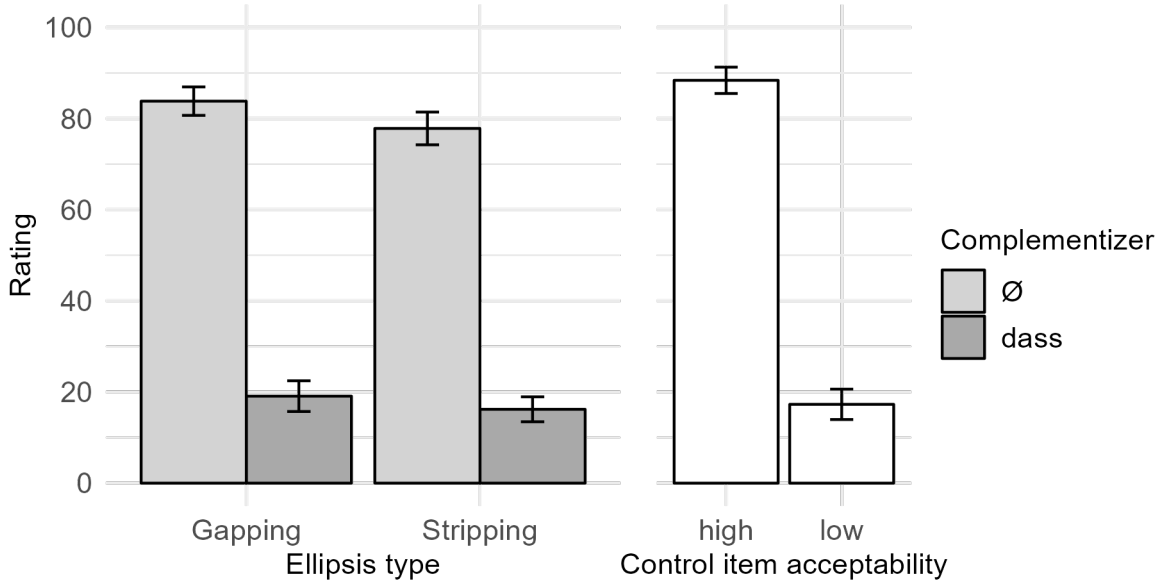


Figure 1: Results for Experiment 1 (gapping vs. stripping under non-factives): Mean ratings and 95% confidence intervals.

0.001). The effect of the ellipsis type was significant as well ($b = 2.11, SE = 0.81, t = 2.60, p = 0.02$). The interaction effect was not significant ($b = 0.87, SE = 0.88, t = 0.99, p = 0.33$).

3 Experiment 2: Factives

Experiment 2 tested gapping and stripping under emotive verbs (i.e. factives). In (7), there is a sample item with the embedding verb *missbilligt* 'disapproves' in all four conditions (same as the sentence in 5 above).

(7) **Target sentence Experiment 2**
Sarah missbilligt, dass der König jeden Tag raucht und {dass / Ø} die Königin {manchmal / auch}.
Sarah disapproves that the king every day smokes and that the queen sometimes too
'Sarah disapproves of the fact that the king smokes every day and that the queen smokes sometimes / that the queen smokes every day, too.'

42 participants completed the questionnaire. 30 of them were from Amazon Mechanical Turk (mturk.com), three were from Prolific.com. The remaining nine participants were students of an introductory linguistics course at the University of Cologne, who volunteered to participate in the experiment without reward. The data from 13 bilingual participants did not enter the analysis. Three further participants made no significant difference between acceptable and non-acceptable controls (as determined by a one-sided t-test per participant) and/or answered more than 20% of the verification questions to the control items incorrectly. The data from these participants similarly did not enter the analysis. This leaves the data from 27 participants in the data set.

The results are in Table 3 and Figure 2. The conditions with *dass* receive very low ratings, which are virtually indistinguishable from the ratings for the unacceptable control items. The conditions without *dass* receive much higher ratings, which approach the ratings of the acceptable controls. Without *dass*, the mean rating for stripping is slightly lower than that of gapping. The results are thus very similar to those of Experiment 1.

Table 3: Experiment 2 (gapping vs. stripping under factives): Mean ratings per experimental condition and control items (*N* = number of observations, *SD* = standard deviation, *SE* = standard error, *CI* = 95% confidence interval)

Ellipsis Type	Complementizer	N	Mean Rating	SD	SE	CI
Gapping	*dass*	160	27.00	24.29	1.92	3.79
Gapping	Ø	157	79.51	27.60	2.20	4.35
Stripping	*dass*	160	25.34	25.30	2.00	3.95
Stripping	Ø	161	68.50	28.51	2.25	4.44
High acceptability controls		480	90.12	26.30	1.20	2.36
Low acceptability controls		465	25.96	39.79	1.85	3.63

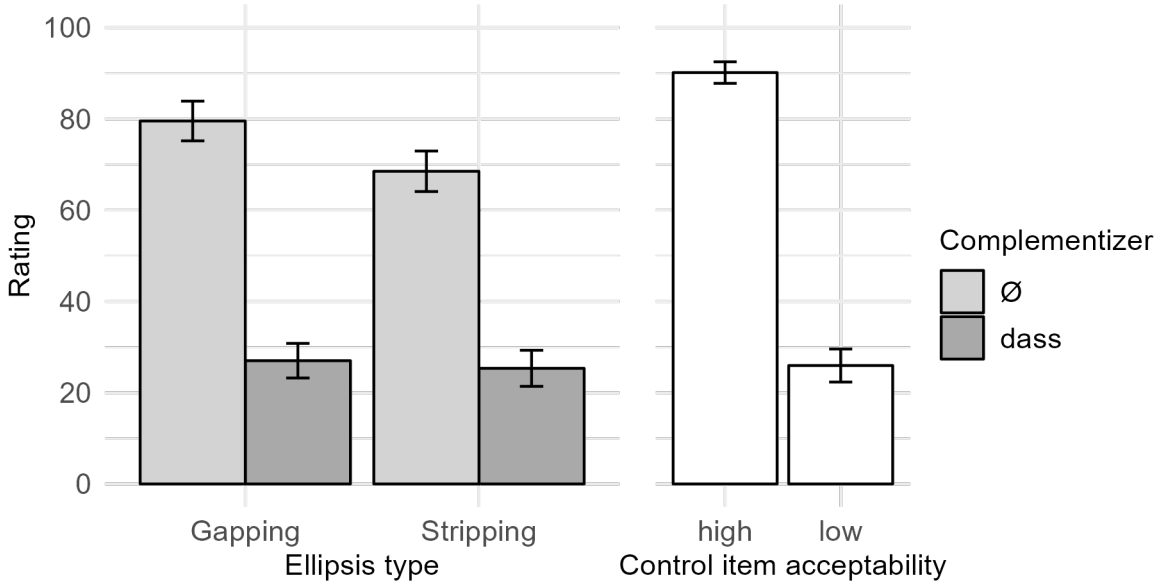


Figure 2: Results for Experiment 2 (gapping vs. stripping under factives): Mean ratings and 95% confidence intervals.

Like for Experiment 1, I think the comparison of the experimental conditions to the acceptable and unacceptable controls suffices to illuminate the pattern of *dass* in gapping and stripping. I nevertheless report the results of a statistical model for sake of completeness. The model indicated the expected significant effect of *dass* ($b = 23.88, SE = 2.52, t = 9.48, p < 0.001$). The effect of the ellipsis type was significant as well ($b = 3.12, SE = 1.20, t = 2.59, p = 0.02$). The interaction effect was not significant ($b = 2.04, SE = 1.33, t = 1.53, p = 0.14$). In terms of statistical significance, the model for Experiment 2 thus parallels the model of Experiment 1.

4 Discussion and conclusion

In both experiments, the mean ratings for gapping and stripping with the complementizer *dass* are virtually indistinguishable from the unacceptable control items. These results straightforwardly show that *dass* is unacceptable in German gapping and stripping. The introspective judgments in the theoretic literature on German that claim that *dass* is ungrammatical thus pass the "sanity check". This observation on the complementizer is the central finding of this chapter, because it relates directly to the topic of this dissertation. In the following, I will touch upon some patterns I observed in the data from these two experiments, which are not central to this dissertation, but nevertheless noticeable and worth mentioning.

The results of both experiments show that gapping and stripping receive high ratings without *dass*, which indicates that these two ellipsis types are acceptable without the complementizer in German. Noticeably though, the relevant conditions do not receive ratings that are quite as high as those for the acceptable controls. I hypothesize that this pattern is an effect of the complexity of the experimental items. The sentences in the experimental materials were structurally complex, containing the three syntactic features clausal coordination, clausal subordination, and ellipsis at the same time. While the acceptable control items were generally similar to the test items, they each contained at most one of these syntactic features, and thus were not quite as complex, perhaps leading to the slight degradation we observe in the results.

Furthermore, both experiments show a significant effect of the ellipsis type, with gapping receiving higher ratings than stripping, most notably in the acceptable conditions without *dass*. I suspect that this effect is independent of embedding. I speculate that it may be the result of a redundancy effect triggered by the polarity particle *auch* 'too'. A stripping sentence with *auch* like the one in (8) might be generally dispreferred to a semantically equivalent sentence like (9), which involves coordinate subjects and no clausal coordination. While not outright unacceptable, stripping with *auch* might receive lower ratings due to the straightforward availability of this structurally simpler alternative.

(8) **Matrix *auch*-stripping** (non-embedded equivalent of the sample item for Experiment 1 in 4/6)
Der König raucht jeden Tag raucht und die Königin auch.
the king smokes every day smokes and the queen too
'The king smokes every day and the queen smokes every day, too.'

(9) **Coordinate subjects**
Der König und die Königin rauchen jeden Tag.
the king and the queen smoke every day
'The king and the queen smoke every day.'

In Bonke (forthcoming) I find a similar effect of stripping with *tože* 'too' in Russian, where I make an equivalent speculation to account for the effect. See also the data for Russian non-embedded coordinate clauses in Appendix B, which also show the *tože*-effect. For Spanish however, sentences with stripping and *también* 'too', seem to be just as acceptable as sentences with gapping (Experiments 1 and 2 in Chapter 3). Spanish thus seems to pattern differently than Russian and German in being more tolerant of (potentially) redundant clausal coordination with *también* 'too'. I leave a closer cross-linguistic examination of this effect to future research.

To reiterate the most important finding, this chapter has provided quantitative evidence in favour of the introspective judgments in the literature on German gapping. The complementizer *dass* indeed leads to unacceptability. The results of two experiments have shown that the effect of *dass* does not result in a mere degradation in acceptability, but outright unacceptability. The same pattern applies to stripping with *auch* 'too'. The comparison of the two experiments shows that the type of embedding verb has no influence on the acceptability of *dass* (non-factives in Experiment 1 vs. factives in Experiment 2). Thus, if we compare the German data of this chapter to the Spanish data of Chapter 3, we observe that the patterns in German and Spanish are genuinely different.

Appendix B: The Russian coordinators *i* and *a* in non-embedded gapping and stripping with *tože* 'too'

In Chapter 2.4, I report the acceptability judgment in (1) on the difference between the coordinating conjunctions *i* and *a*, both 'and', in gapping. My claim was that while *a* is fully acceptable and thus the natural choice of coordinator, *i* is degraded but not outright unacceptable. I describe this as a "soft constraint" on the coordinator in Chapter 2.4. The present chapter provides supporting evidence for this acceptability judgment.

(1) Vasja poedet v Moskvu, {a / ??i } Petja v Peterburg.
Vasja will.go to Moscow and and Pete to St. Petersburg
'Vasja will go to Moscow and Pete to St. Petersburg.' (reproduced from example 103 in Chapter 2.4)

I conducted an acceptability experiment that served as a pretest for the experimental items in Bonke (forthcoming). To see whether the lexical contexts of the items were overall acceptable independently of clausal embedding, the experimental materials contained the items from Bonke (forthcoming), but without clausal embedding. Items came in two ellipsis types, gapping and stripping with *tože*. I also wanted to check whether my prospective choices for the coordinator with these two ellipsis types (*a* for gapping, *i* for *tože*-stripping) were correct. This resulted in a 2 ×2 design with the factors ellipsis (gapping vs. *tože*-stripping) and coordinator (*a* vs. *i*).

The resulting four conditions are illustrated for a sample item in (2) and (3). Notice that the lexical content in that example is a translation equivalent of the embedded coordinate clauses in the Spanish sample items in Chapter 3.3 and the German sample items in Appendix A. This was a deliberate choice. In total, the present pretest experiment featured 24 such lexical contents, all of which had translation equivalents in the Spanish/German examples. However, some of these lexical contents received overall low acceptability ratings and were replaced in the materials for Bonke (forthcoming). I will not go into detail of these individual lexical contents since this discussion would be besides the point of this chapter. The sentence in (1) above did not feature in the experimental materials; I chose it in Chapter 2.4 to illustrate the coordinator patterns based on an example from the literature that had previously been introduced in the text.

(2) **Sample item, gapping conditions**
Korol' kurit každyj den', {a / i } koroleva – inogda.
king smokes every day and and queen sometimes.
'The king smokes every day and the queen sometimes.'

(3) **Sample item, *tože*-tripping conditions**

Korol' kurit každyj den', {a / i } koroleva – tože.
king smokes every day and and queen too
'The king smokes every day and the queen, too.'

The items were distributed across four lists in a Latin-square design. Each list featured each of the four the items with locally embedded gapping from Chapter 2.4 in one of the four conditions as filler items. (This means that the experiment discussed in Chapter 2.4 is identical to the present one.) For details on these filler items, as well as a description of the procedure and participants, see that chapter. I will go straight to the discussion of the results, which are in Table 1 and Figure 1.

Table 1: Mean ratings per experimental condition and unacceptable controls (*N* = number of observations, *SD* = standard deviation, *SE* = standard error, *CI* = 95% confidence interval)

Ellipsis	Coordinator	N	Mean rating	SD	SE	CI
Gapping	*a*	216	88.87	17.45	1.19	2.34
Gapping	*i*	216	38.53	26.53	1.81	3.56
tože-Stripping	*a*	216	26.72	23.30	1.59	3.12
tože-Stripping	*i*	216	75.17	24.40	1.66	3.27
Unacceptable controls		108	19.64	22.95	2.21	4.38

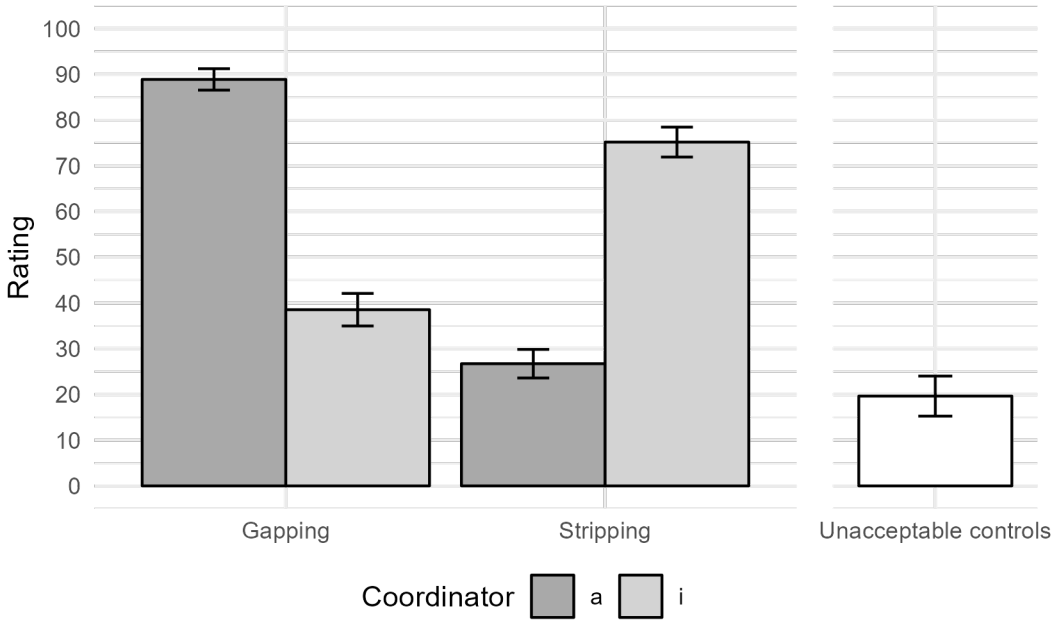


Figure 1: Acceptability of the experimental conditions and unacceptable controls. Mean ratings with 95% confidence interval

As expected, *a* is the appropriate choice for gapping, and *i* is the appropriate choice for *tože*-stripping. These two combinations receive considerably higher ratings than their respective alternatives. For gapping, this alternative is the coordinator *i*, which is the crucial condition for the judgment in (1). While this configuration (gapping with *i*) is strongly degraded in comparison to its variant with *a*, it does not seem

to be outright unacceptable. There is still a noticeable difference between this condition and the unacceptable controls. This result justifies the judgment in (1), and the characterization of coordinator choice in gapping as a soft constraint. However, more research is needed to fully illuminate the constraints on coordinator choice.

Finally, notice that stripping with *tože* received lower ratings than gapping, for both the better rated coordinator (*i* for *tože*-stripping, *a* for gapping) and the worse rated coordinator (*a* for *tože*-stripping, *i* for gapping). To investigate this apparent difference between the ellipsis types, I used `R` (R Core Team 2023) and the package `afex` (Singmann et al. 2022) to fit a linear mixed effects model to the data. The model contained the coordinator (*i* vs. *a*) and the ellipsis type (gapping vs. *tože*-stripping) plus the interaction between the two as fixed factors. The random effects structure was maximal (i.e. random intercepts and slopes for items and participants for the fixed factors and the interaction). Factors were sum-to-zero contrast coded. The model indicated a statistically significant effect of the type of ellipsis ($b = -6.38, SE = 1.13, t = -5.65, p < 0.001$).[3] This statistical result suggests that stripping with *tože* is indeed less acceptable than gapping. In Bonke (forthcoming) I discuss similar findings for Russian globally embedded gapping and *tože*-stripping. Furthermore, in Appendix A I discuss similar findings for German gapping and stripping with *auch* 'too'. See that paper and/or that appendix for a brief speculative analysis in terms of a redundancy effect caused by *tože*/*auch*.

[3]The effect of the coordinator was not significant ($b = -0.47, SE = 1.37, t = -0.35, p = 0.73$). As one would expect from a visual inspection of Figure 1, the interaction effect was large and significant ($b = 24.70, SE = 2.00, t = 12.36, p < 0.001$).

References

Abe, Jun & Hiroto Hoshi. 1997. Gapping and P-Stranding. *Journal of East Asian Linguistics* 6(1). 101–136. DOI: 10.1023/A:1008240005018.

Abeillé, Anne, Gabriela Bîlbîie & Francois Mouret. 2014. A Romance perspective on gapping constructions. In *Romance Perspectives on Construction Grammar*. Hans C. Boas & Francisco Gonzálvez-García (eds.). Amsterdam: John Benjamins. 227–265. DOI: 10.1075/cal.15.07abe.

Abrusán, Márta. 2011. Predicting the presuppositions of soft triggers. *Linguistics and Philosophy* 6. 491–535. DOI: 10.1007/s10988-012-9108-y.

Alves Castro, Marian. 2016. Gapping and Pseudogapping Constructions with Multiple Remnants. In *On the Move: Glancing Backwards To Build a Future in English Studies*. Aitor Ibarrola-Armendariz & Jon Ortiz de Urbina Arruabarrena (eds.). Bilbao: Universidad de Deusto. 173–181.

Anand, Pranav & Valentine Hacquard. 2014. Factivity, belief and discourse. In *The art and craft of semantics: A festschrift for Irene Heim*. Luka Crnič & Uli Sauerland (eds.). Cambridge, MA: MIT Working Papers in Linguistics. 69–90.

Bailyn, John Frederick & Tatiana Bondarenko. 2018. Russian. In Jeroen van Cranenbroek & Tanja Temmerman (eds.), *The Oxford Handbook of Ellipsis*. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780198712398.013.36.

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278. DOI: 10.1016/j.jml.2012.11.001.

Bassi, Itai & Tatiana Bondarenko. 2021. Composing CPs: Evidence from disjunction and conjunction. In Joseph Rhyne, Kaelyn Lamp, Nicole Dreier & Chloe Kwon (eds.), *Proceedings of the 30th Semantics and Linguistic Theory Conference*, 583–602. DOI: 10.3765/salt.v30i0.4837.

Beaver, David. 2010. Have You Noticed that Your Belly Button Lint Color is Related to the Color of Your Clothing? In *Presuppositions and Discourse: Essays Offered to Hans Kamp*. Rainer Bauerle, Uwe Reyle & Thomas Zimmermann (eds.). Leiden: Brill. 65–100. DOI: 10.1163/9789004253162_004.

Bîlbîie, Gabriela & Israel de la Fuente. 2019. Can gapping be embedded? Experimental evidence from Spanish. *Glossa* 4(1). 110. DOI: 10.5334/gjgl.782.

Bîlbîie, Gabriela, Israel de la Fuente & Anne Abeillé. 2023. Factivity and complementizer omission in English embedded gapping. *Journal of Linguistics* 59(2). 389–426. DOI: 10.1017/S0022226722000445.

Bîlbîie, Gabriela, Israel de la Fuente & Anne Abeillé. 2021. L'ellipse verbale enchâssée dans les langues romanes. *Langages* 223(3). 61–80. http://www.revues.armand-colin.com/lettres-langues/langages/langages-no-223-32021/lellipse-verbale-enchassee-langues-romanes.

Bîlbîie, Gabriela & Pegah Faghiri. 2022. An experimental perspective on embedded gapping in Persian. *The Linguistic Review* 39(3). 557–586. DOI: 10.1515/tlr-2022-2097.

Bolinger, Dwight. 1968. Postposed main phrases: an English rule for the Romance subjunctive. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 14(1). 3–30. DOI: 10.1017/S0008413100019629.

Bolinger, Dwight. 1972. *That's that*. The Hague: Mouton.

Bonke, Max. Forthcoming. The complementizer *čto* in Russian gapping and stripping. *Proceedings of FDSL 16*.

Bonke, Max & Sophie Repp. 2022. Complementizer deletion in embedded gapping in Spanish. *The Linguistic Review* 39(3). 525–555. DOI: 10.1515/tlr-2022-2096.

Boone, Enrico. 2014. *The syntax and licensing of Gapping and Fragments*. Utrecht: LOT.

Borrego Nieto, Julio, José Jésus Gómez Asencio & Emilio Prieto de los Mozos. 1986. *El subjuntivo: valores y usos*. Madrid: Sociedad General Española de Libreria.

Bošković, Željko & Howard Lasnik. 2003. On the distribution of null complementizers. *Linguistic Inquiry* 34(4). 527–546. http://www.jstor.org/stable/4179250.

Broekhuis, Hans & Norbert Corver. 2019. *Syntax of Dutch: Coordination and Ellipsis*. Amsterdam: Amsterdam University Press.

Brovetto, Claudia. 2002. Spanish clauses without complementizer. In *Current Issues in Romance Languages: Selected papers from the 29th Linguistic Symposium on Romance Languages (LSRL), Ann Arbor, 8–11 April 1999*. Teresa Satterfield, Christina Tortora & Diana Cresti (eds.). John Benjamins. 33–46.

Centeno, Naiara. 2011. *Gapping and determiner sharing in Spanish*. Euskal Herriko Unibertsitatea-University of the Basque Country. (Doctoral dissertation).

Chierchia, Gennaro & Sally McConnell-Ginet. 1996. *Meaning and grammar*. 5th ed. Cambridge, MA: MIT Press.

Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT Press.

Coppock, Elizabeth. 2001. Gapping: in defense of deletion. In Mary Andronis, Christopher Ball, Heidi Elston & Sylvain Neuvel (eds.), *Proceedings from the Annual Meeting of the Chicago Linguistic Society 37*, 133–148. Chicago: Chicago Linguistic Society.

Corr, Alice. 2018. 'Exclamative' and 'quotative' illocutionary complementisers in Catalan, European Portuguese and Spanish: A study in Ibero-Romance syntactic 'near-synonymy'. *Languages in Contrast* 18(1). 69–98. DOI: 10.1075/lic.00004.cor.

Corr, Alice. 2022. *The Grammar of the Utterance: How to Do Things with Ibero-Romance*. Oxford: Oxford University Press. DOI: 10.1093/oso/9780198856597.001.0001.

de Cuba, Carlos & Jonathan E. MacDonald. 2013. On the referential status of embedded polarity answers in Spanish. In Jennifer Cabrelli Amaro, Gillian Lord, Ana de Prada Pérez & Jessi Elana Aaron (eds.), *Selected proceedings of the 16th Hispanic Linguistics Symposium*, 312–323. Somerville, MA: Cascadilla.

Degen, Judith & Judith Tonhauser. 2021. Prior Beliefs Modulate Projection. *Open Mind* 5. 59–70. DOI: 10.1162/opmi_a_00042.

Degen, Judith & Judith Tonhauser. 2022. Are there factive predicates? An empirical investigation. *Language* 98(3). 552–591. DOI: 10.1353/lan.0.0271.

Delbecque, Nicole & Béatrice Lamiroy. 2000. La subordinación sustantiva: Las subordinadas enunciativas en los complementos verbales. In *Gramática descriptiva de la lengua española*. Ignacio Bosque & Violeta Demonte (eds.). 1. ed., 3. reprint. Vol. 2. Madrid: Espasa. 1965–2082.

Demonte, Violeta & Olga Fernández-Soriano. 2009. Force and finiteness in the Spanish complementizer system. *Probus* 21(1). 23–49. DOI: 10.1515/prbs.2009.002.

Depiante, Marcela A. 2004. Dos casos de elipsis con particula de polaridad en Español - Evidencia a favor de una visión no uniforme de la elipsis. *RASAL Lingüistica* 1. 53–69.

Djärv, Kajsa & Hezekiah Akiva Bacovcin. 2017. Prosodic effects on factive presupposition projection. In Dan Burgdorf, Jacob Collard, Sireemas Maspong & Brynhildur Stefánsdóttir (eds.), *Proceedings of SALT 27*, 116–133. DOI: 10.3765/salt.v27i0.4134.

Dor, Daniel. 2005. Toward a semantic account of *that*-deletion in English. *Linguistics* 43(2). 345–382. DOI: 10.1515/ling.2005.43.2.345.

Egré, Paul. 2008. Question-embedding and factivity. *Grazer Philosophische Studien* 77(1). 85–125. DOI: 10.1163/18756735-90000845.

Elsness, Johan. 1984. *That* or zero? A look at the choice of object clause connective in a corpus of American English. *English Studies* 65(6). 519–533. DOI: 10.1080/00138388408598357.

Erschler, David. 2018. *Typology of bizarre ellipsis varieties*. University of Massachusetts Amherst. (Doctoral dissertation). DOI: 10.7275/12727484.

Erteschik-Shir, Nomi. 1977. *On the nature of island constraints*. Bloomington, Indiana: Indiana Univ. Linguistics Club.

Etxepare, Ricardo. 2010. From hearsay evidentiality to samesaying relations. *Lingua* 120(3). 604–627. DOI: 10.1016/j.lingua.2008.07.009.

Fábregas, Antonio. 2014. A guide to subjunctive and modals in Spanish: questions and analyses. *Borealis* 3(2). 1–94. DOI: 10.7557/1.3.2.3064.

Farkas, Donka. 2003. Assertion, belief and mood choice. Paper presented at ESSLLI, Conditional and Unconditional Modality Workshop. Vienna. https://people.ucsc.edu/%7Efarkas/papers/mood.pdf.

Farkas, Donka F. & Kim B. Bruce. 2009. On Reacting to Assertions and Polar Questions. *Journal of Semantics* 27(1). 81–118. DOI: 10.1093/jos/ffp010.

Farudi, Annahita. 2013. *Gapping in Farsi: A crosslinguistic investigation*. University of Massachusetts Amherst. (Doctoral dissertation).

Faulkner, Tris. 2021a. *A systematic investigation of the Spanish subjunctive: Mood variation in subjunctive clauses*. Georgetown University. (Doctoral dissertation).

Faulkner, Tris. 2021b. Prescriptively or descriptively speaking?: How 'information-quality' influences mood variation in Spanish emotive-factive clauses. *Pragmatics* 31(3). 357–381. DOI: 10.1075/prag.19044.fau.

Fernández-Sánchez, Javier. 2019. Against a clausal ellipsis account of all stripping strings in spanish. *Poznan Studies in Contemporary Linguistics* 55(1). 53–87. DOI: 10.1515/psicl-2019-0003.

Fernández-Sánchez, Javier. 2021. On complementizers and embedded gapping in English, Spanish and Polish. *Beyond Philology* (3). 109–135. DOI: 10.26881/bp.2021.3.04.

Ferreira, Victor S. & Melanie Hudson. 2011. Saying "that" in dialogue: the influence of accessibility and social factors on syntactic production. *Language and Cognitive Processes* 26(10). 1736–1762. DOI: 10.1080/01690965.2010.537482.

Féry, Caroline & Katharina Hartmann. 2005. The focus and prosodic structure of German Right Node Raising and Gapping. *The Linguistic Review* 22(1). 69–116. DOI: 10.1515/tlir.2005.22.1.69.

Fiengo, Robert W. 1974. *Semantic conditions on surface structures*. MIT. (Doctoral dissertation). http://hdl.handle.net/1721.1/13018.

García Núñez, José María. 2020. On the left periphery of spanish indirect interrogatives. *Probus* 32(1). 55–92. DOI: 10.1515/probus-2019-0005.

García Núñez, José María. 2021. Embedded root phenomena and indirect speech reports. *Linguistics* 59(6). 1531–1566. DOI: 10.1515/ling-2021-0162.

Garcia-Marchena, Oscar. 2018. Polar Verbless Clauses and Gapping Subordination in Spanish. In *Grammar and Corpora 2016*. Eric Fuß, Marek Konopka, Beata Trawiński & Ulrich H. Waßner (eds.). Heidelberg University Publishing. DOI: 10.17885/heiup.361.c4697.

Gazdar, Gerald. 1979. *Pragmatics: implicature, presupposition, and logical form*. New York: Academic Press.

Giannakidou, Anastasia. 1998. *Polarity Sensitivity as (Non)Veridical Dependency*. Amsterdam: John Benjamins. DOI: 10.1075/la.23.

Giannakidou, Anastasia & Alda Mari. 2019. Emotive predicates and the subjunctive: a flexible mood OT account based on (non)veridicality. In Nadine Bade, Polina Berezovskaya & Anthea Schöller (eds.), *Proceedings of Sinn und Bedeutung 20*, 288–305. https://ojs.ub.uni-konstanz.de/sub/index.php/sub/article/view/264.

Gildersleeve, Basil L. & Gonzalez Lodge. 1905. *Gildersleeve's Latin grammar*. New York, Boston: University Publishing Company. http://hdl.handle.net/2027/mdp.39015002651332.

Ginzburg, Jonathan & Ivan A. Sag. 2001. *Interrogative investigations*. Stanford, CA: CSLI.

Grebenyova, Lydia. 2012. *Syntax, Semantics and Acquisition of Multiple Interrogatives: Who wants what?* Amsterdam: John Benjamins. DOI: 10.1075/la.195.

Greenberg, Yael & Lavi Wolf. 2018. Gradable assertion speech acts. In Sherry Hucklebridge & Max Nelson (eds.), *Proceedings of NELS 48*, 271–280.

Haegeman, Liliane. 2006. Conditionals, factives and the left periphery. *Lingua* 116(10). 1651–1669. DOI: 10.1016/j.lingua.2005.03.014.

Haegeman, Liliane & Terje Lohndal. 2015. Be careful how you use the left periphery. In Elisa Di Domenico, Cornelia Hamann & Simona Matteini (eds.), *Structures, strategies and beyond: studies in honour of adriana belletti*, vol. 223 (Linguistik Aktuell/Linguistics Today), 135–162. Amsterdam John Benjamins.

Hankamer, Jorge. 1979. *Deletion in Coordinate Structures*. New York & London: Garland.

Hansen, Björn, Alexander Letuchiy & Izabela Błaszczyk. 2016. Complementizers in Slavonic (Russian, Polish, and Bulgarian). In *Complementizer Semantics in European Languages*. Kasper Boye & Petar Kehayov (eds.). Berlin, Boston: De Gruyter Mouton. 175–224. DOI: 10.1515/9783110416619-008.

Hartmann, Katharina. 2000. *Right Node Raising and Gapping: Interface conditions on prosodic deletion*. Amsterdam: John Benjamins. DOI: 10.1075/z.106.

Heim, Irene. 1992. Presupposition Projection and the Semantics of Attitude Verbs. *Journal of Semantics* 9(3). 183–221. DOI: 10.1093/jos/9.3.183.

Hooper, Joan B. 1975. On Assertive Predicates. In *Syntax and Semantics volume 4*. John P. Kimball (ed.). Leiden: Brill. 91–124. DOI: 10.1163/9789004368828_005.

Hooper, Joan B. & Sandra A. Thompson. 1973. On the applicability of root transformations. *Linguistic Inquiry* 4(4). 465–497. http://www.jstor.org/stable/4177789.

Hope, Ryan M. 2022. *Rmisc: Ryan Miscellaneous*. R package version 1.5.1. https://CRAN.R-project.org/package=Rmisc.

Hosking, Richard. 1996. *A dictionary of Japanese food*. Boston, Rutland, Tokyo: Tuttle.

Hoyt, Frederick & Alexandra Teodorescu. 2012. How many kinds of sluicing, and why? Single and multiple sluicing in Romanian, English, and Japanese. In Jason Marchant & Andrew Simpson (eds.), *Sluicing: Cross-Linguistic Perspectives*, 83–103. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199645763.003.0005.

Ichikawa, Jonathan Jenkins & Matthias Steup. 2018. The Analysis of Knowledge. In Edward N. Zalta (ed.), *The Stanford encyclopedia of philosophy*. Stanford: Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2018/entries/knowledge-analysis.

İnce, Atakan. 2009. *Dimensions of Ellipsis: Investigations in Turkish*. University of Maryland. (Doctoral dissertation). http://hdl.handle.net/1903/9838.

Izutsu, Mitsuko Narita. 2008. Contrast, concessive, and corrective: Toward a comprehensive study of opposition relations. *Journal of Pragmatics* 40(4). 646–675. DOI: 10.1016/j.pragma.2007.07.001.

Jackendoff, Ray S. 1971. Gapping and related rules. *Linguistic Inquiry* 2(1). 21–35. http://www.jstor.org/stable/4177607.

Jaeger, Florian T. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology* 61(1). 23–62. DOI: 10.1016/j.cogpsych.2010.02.002.

Jasinskaja, Katja & Henk Zeevat. 2008. Explaining additive, adversative and contrast marking in Russian and English. *Revue de Sémantique et Pragmatique* 24(1). 65–91.

Jasinskaja, Katja & Henk Zeevat. 2009. Explaining conjunction systems: Russian, english, german. In Arndt Riester & Torgrim Solstad (eds.), *Proceedings of Sinn und Bedeutung 13*, 231–246.

Johnson, Kyle. 2009. Gapping Is Not (VP-) Ellipsis. *Linguistic Inquiry* 40(2). 289–328. DOI: 10.1162/ling.2009.40.2.289.

Johnson, Kyle. 2014. Gapping. Unpublished manuscript. University of Massachusetts, Amherst.

Johnson, Kyle. 2018. Gapping and Stripping. In Jeroen van Cranenbroek & Tanja Temmerman (eds.), *The Oxford Handbook of Ellipsis*. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780198712398.013.24.

Johnson, Kyle. 1996/2004. In search of the English middle field. Unpublished manuscript. University of Massachusetts, Amherst.

Jung, Wonsuk. 2017. *The non-unity of gapping*. Euskal Herriko Unibertsitatea-University of the Basque Country. (Doctoral dissertation).

Jung, Wonsuk. 2023. The syntax of two types of gapping in Spanish. *Studia Linguistica* 77(1). 77–121. DOI: 10.1111/stul.12204.

Kalinin, Aleksandr. 2020. Gapping is not only low coordination. *Glossa* 5(1). 46. DOI: 10.5334/gjgl.1044.

Karttunen, Lauri. 1971. Some observations on factivity. *Papers in Linguistics* 4. 55–69. DOI: 10.1080/08351817109370248.

Karttunen, Lauri. 2016. Presupposition: What went wrong? In Mary Moroney, Carol-Rose Little, Jacob Collard & Dan Burgdorf (eds.), *Proceedings of salt 26*, 705–731. DOI: 10.3765/salt.v26i0.3954.

Kastner, Itamar. 2015. Factivity mirrors interpretation: the selectional requirements of presuppositional verbs. *Lingua* 164. 156–188. DOI: 10.1016/j.lingua.2015.06.004.

Kazenin, Konstantin. 2010. Russian gapping: Against ATB. In Gerhild Zybatow, Philip Dudchuk, Serge Minor & Ekaterina Pshehotskaya (eds.), *Proceedings of formal description of Slavic languages 7.5*, 85–100. Frankfurt am Main: Peter Lang.

Kempchinsky, Paula. 1998. Mood Phrase, Case Checking and Obviation. In *Romance Linguistics: Theoretical Perspectives. Selected papers from the 27th Linguistic Symposium on Romance Languages (LSRL XXVII), Irvine, 20–22 February, 1997*. Armin Schwegler, Bernard Tranel & Myriam Uribe-Etxebarria (eds.). Amsterdam: John Benjamins. 143–154. https://www.jbe-platform.com/content/books/9789027275691.

King, Tracy H. 1993. *Configuring topic and focus in Russian*. Stanford University. (Doctoral dissertation). https://www.proquest.com/dissertations-theses/configuring-topic-focus-russian/docview/304080174/se-2.

Kiparsky, Paul & Carol Kiparsky. 1970. Fact. In *Progress in Linguistics*. Manfred Bierwisch & Karl Erich Heidolph (eds.). Berlin, Boston: De Gruyter Mouton. 143–173. DOI: 10.1515/9783111350219.143.

Klein, Ewan. 1975. Two sorts of factive predicate. *Pragmatic Microfiche* 1(1). B5–C14.

Klein, Philip W. 1977. Semantic Factors in Spanish Mood. *Glossa: An International Journal of Linguistics* 11. 3–19.

Knyazev, Mikhail. 2022. Spanning complement-taking verbs and spanning complementizers: On the realization of presuppositional clauses. *Journal of Linguistics* 58(4). 887–899. DOI: 10 . 1017 / S0022226722000160.

Kuno, Susumu. 1976. Gapping: A Functional Analysis. *Linguistic Inquiry* 7(2). 300–318. http://www.jstor.org/stable/4177923.

Kush, Dave. 2016. Notes on gapping in Hindi-Urdu. *Linguistic Analysis* 40(3–4). 255–296.

Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software* 82(13). 1–26. DOI: 10.18637/jss.v082.i13.

Lahiri, Utpal. 2002. *Questions and Answers in Embedded Contexts*. Oxford: Oxford University Press. DOI: 10.1093/oso/9780198241331.001.0001.

Laskurain–Ibarluzea, Patxi. 2015. Quantification and Mood Distribution in Spanish Complements: On the Negative Features of poco/a/s in Spanish. *Hispania* 98(2). 285–299. http://www.jstor.org/stable/24368288.

Leiner, Dominic. 2024. *SoSci Survey*. Version 3.5.01. https:/www.soscisurvey.de.

Lin, Vivian I-Wen. 2002. *Coordination and sharing at the interfaces*. MIT. (Doctoral dissertation). http://hdl.handle.net/1721.1/8151.

Lleó, Conxita. 1979. *Some optional rules in Spanish complementation*. Tübingen: Max Niemeyer.

Lope Blanch, Juan M. 1958. Algunos usos de indicativo por subjuntivo en oraciones subordinadas. *Nueva Revista de Filología Hispánica* 12(3/4). 383–385. http://www.jstor.org/stable/40297333.

López, Luis & Susanne Winkler. 2003. Variation at the syntax-semantics interface: Evidence from gapping. In Kerstin Schwabe & Susanne Winkler (eds.), *The Interfaces: Deriving and interpreting omitted structures*, 227–248. Amsterdam/Philadelphia: John Benjamins. https://www.jbe-platform.com/content/books/9789027296917-la.61.11lop.

Lu, Jiayi & Nayoun Kim. 2022. The puzzle of argument structure mismatch in gapping. *Frontiers in Psychology* 13. DOI: 10.3389/fpsyg.2022.907823.

Lunn, Patricia V. 1989. Spanish mood and the prototype of assertability. *Linguistics* 27(4). 687–702. DOI: 10.1515/ling.1989.27.4.687.

Martínez Vera, Gabriel. 2020. On recomplementation, high adverbs and V-movement in Spanish. In Irene Vogel (ed.), *Romance Languages and Linguistic Theory 16: Selected papers from the 47th Linguistic Symposium on Romance Languages (LSRL), Newark, Delaware*, 187–202. Amsterdam: John Benjamins. DOI: 10.1075/rllt.16.12mar.

Mazzola, Giulia. 2022. *Syndetic and asyndetic complementation in Spanish*. KU Leuven. (Doctoral dissertation).

Meinunger, André. 2017. Desires, wishes and hopes – desiderative predicates and presuppositions. *STUF - Language Typology and Universals* 70(4). 611–635. DOI: 10.1515/stuf-2017-0026.

Mejías-Bikandi, Errapel. 1994. Assertion and Speaker's Intention: A Pragmatically Based Account of Mood in Spanish. *Hispania* 77(4). 892–902. http://www.jstor.org/stable/345752.

Mendoza, Imke. 1996. *Zur Koordination im Russischen*. München: Sagner. https://mdz-nbn-resolving.de/bsb00051910.

Merchant, Jason. 2001. *The Syntax of Silence: Sluicing, Islands, and the Theory of Ellipsis*. Oxford: Oxford University Press. DOI: 10.1093/oso/9780199243730.001.0001.

Messick, Troy & Deepak Alok. 2021. Stripping in Hindi: Does clause size matter? In *The size of things I: Structure building*. Sabine Laszakovits & Zheng Shen (eds.). Berlin: Language Science Press. 247–263. DOI: 10.5281/zenodo.5524295.

Morgan, Jerry L. 1973. Sentence fragments and the notion "sentence." In *Issues in linguistics: Papers in honor of Henry and Renée Kahane*. Braj B. Kachru, Robert Lees, Yakov Malkiel, Angelina Pietrangeli & Sol Saporta (eds.). Urbana-Champaign: Univ. of Illinois Press. 719–751.

Nunes, Jairo. 2001. Sideward Movement. *Linguistic Inquiry* 32(2). 303–344. DOI: 10.1162/00243890152001780.

Nunes, Jairo. 2004. *Linearization of chains and sideward movement*. Cambridge, MA: MIT Press.

Ott, Dennis & Volker Struckmeier. 2018. Particles and Deletion. *Linguistic Inquiry* 2(49). 393–407. DOI: 10.1162/LING_a_00277.

Paredes García, Florentino. 2023. Queísmo y dequeísmo en español: patrones sociolingüísticos y geolectales. *Círculo de Lingüística Aplicada a la Comunicación* 94. 51–63. DOI: 10.5209/clac.84706.

Park, Sang-Hee. 2016. Towards a QUD-Based Analysis of Gapping Constructions. In *Proceedings of the 30th Pacific Asia Conference on Language, Information and Computation: Oral Papers*, 297–305. Seoul, South Korea. https://aclanthology.org/Y16-2028.

Park, Sang-Hee. 2019. *Gapping: A Constraint-Based Syntax-Semantics Interface*. State University of New York at Buffalo. (Doctoral dissertation).

Patin, Cédric & Sophie Manus. 2018. Kiswahili and Shingazidja. In Jeroen van Craenenbroeck & Tanja Temmerman (eds.), *The Oxford Handbook of Ellipsis*, 900–933. Oxford University Press. DOI: 10.1093/oxfordhb/9780198712398.013.41.

Plann, Susan. 1982. Indirect Questions in Spanish. *Linguistic Inquiry* 13(2). 297–312. http://www.jstor.org/stable/4178276.

Potter, David, Michael Frazier & Masaya Yoshida. 2017. A two-source hypothesis for gapping. *Natural Language & Linguistic Theory* 35(4). 1123–1160. DOI: 10.1007/s11049-017-9359-y.

Quer, Josep. 1998. *Mood at the interface*. The Hague: Holland Academic Graphics.

Quer, Josep. 2009. Twists of mood: The distribution and interpretation of indicative and subjunctive. *Lingua* 119(12). 1779–1787. DOI: 10.1016/j.lingua.2008.12.003.

Quer, Josep. 2011. On the (un)stability of mood distribution in Romance. In *Modality and Mood in Romance: Modal interpretation, mood selection, and mood alternation*. Martin G. Becker & Eva-Maria Remberger (eds.). Berlin, New York: De Gruyter. 163–180. DOI: 10.1515/9783110234343.2.163.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/.

Reich, Ingo. 2007. Toward a uniform analysis of short answers and gapping. In Kerstin Schwabe & Susanne Winkler (eds.), *On Information Structure, Meaning and Form*, 467–484. Amsterdam: John Benjamins. DOI: 10.1075/la.100.

Repp, Sophie. 2009. *Negation in Gapping*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199543601.001.0001.

Rivero, María-Luisa. 1971. Mood and Presupposition in Spanish. *Foundations of Language* 7(3). 305–336. http://www.jstor.org/stable/25000537.

Rivero, María-Luisa. 1978. Topicalization and Wh Movement in Spanish. *Linguistic Inquiry* 9(3). 513–517. http://www.jstor.org/stable/4178077.

Rivero, María-Luisa. 1980. On Left-Dislocation and Topicalization in Spanish. *Linguistic Inquiry* 11(2). 363–393. http://www.jstor.org/stable/4178165.

Rizzi, Luigi. 1997. The fine structure of the left periphery. In *Elements of grammar*. Liliane Haegeman (ed.). Dordrecht: Kluwer. 281–337.

Roberts, Craige. 1996. Information structure: Towards an integrated formal theory of pragmatics. In *Ohio State University Working Papers in Linguistics (OSUWPL) Volume 49. Papers in Semantics*. Je Hak Yoon & Andreas Kathol (eds.). Columbus, OH: The Ohio State University Department of Linguistics.

Roberts, Craige. 2012. Information structure in discourse: towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5. Reprint of Roberts 1996, 1–69. DOI: 10.3765/sp.5.6.

Roland, Douglas, Frederic Dick & Jeffrey L. Elman. 2007. Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language* 57(3). 348–379. DOI: 10.1016/j.jml.2007.03.002.

Rosenbaum, Peter S. 1967. *The grammar of english predicate complement constructions*. Cambridge MA: M.I.T. Press.

Ross, John R. 1968. *Constraints on variables in syntax*. Bloomington, Indiana: Indiana Univ. Linguistics Club.

Ross, John R. 1970. Gapping and the order of constituents. In *Progress in Linguistics*. Manfred Bierwisch & Karl Erich Heidolph (eds.). Berlin, Boston: De Gruyter Mouton. 249–259. DOI: doi:10.1515/9783111350219.249.

Sag, Ivan A. 1977. *Deletion and logical form*. Bloomington, Indiana: Indiana Univ. Linguistics Club.

Sailor, Craig & Gary Thoms. 2014. On the Non-existence of Non-constituent Coordination and Non-constituent Ellipsis. In *Proceedings of the 31st West Coast Conference on Formal Linguistics*. Robert E. Santana-LaBarge (ed.). Somerville, MA: Cascadilla. 89–97. https://www.lingref.com/cpp/wccfl/31/index.html.

Schlenker, Philippe. 2005. The Lazy Frenchman's Approach to the Subjunctive (Speculations on Reference to Worlds and Semantic Defaults in the Analysis of Mood). In Twan Geerts, Ivo van Ginneken & Haike Jacobs (eds.), *Romance languages and linguistic theory 2003. Selected papers from 'Going Romance'*, 269–309. Amsterdam/Philadelphia: John Benjamins.

Schlenker, Philippe. 2010. Local contexts and local meanings. *Philosophical Studies* 115. 115–142. DOI: 10.1007/s11098-010-9586-0.

Schwarz, Florian (ed.). 2015. *Experimental perspectives on presuppositions*. Heidelberg: Springer.

Shanon, Benny. 1976. On the two kinds of presuppositions in natural language. *Foundations of Language* 14(2). 247–249. http://www.jstor.org/stable/25170057.

Shim, Ji Young & Tabea Ihsane. 2017. A new outlook of complementizers. *Languages* 2(3). DOI: 10.3390/languages2030017.

Siegel, Muffy E. A. 1984. Gapping and interpretation. *Linguistic Inquiry* 15(3). 523–530. http://www.jstor.org/stable/4178400.

Siegel, Muffy E. A. 1987. Compositionality, case, and the scope of auxiliaries. *Linguistics and Philosophy* 10(1). 53–75. http://www.jstor.org/stable/25001262.

Simons, Mandy, David Beaver, Craige Roberts & Judith Tonhauser. 2017. The Best Question: Explaining the Projection Behavior of Factives. *Discourse Processes* 54(3). 187–206. DOI: 10.1080/0163853X.2016.1150660.

Simons, Mandy, Judith Tonhauser, David Beaver & Craige Roberts. 2010. What projects and why. In Nan Li & David Lutz (eds.), *Proceedings of SALT 20*, 309–327. Ithaca, NY: CLC Publications.

Singmann, Henrik, Ben Bolker, Jake Westfall, Frederik Aust & Mattan S. Ben-Shachar. 2022. *afex: Analysis of Factorial Experiments*. R package version 1.2-0. https://CRAN.R-project.org/package=afex.

Smith, E. Allyn & Kathleen Hall. 2011. Projection diversity: Experimental evidence. In Gregory Kierstead (ed.), *Proceedings of ESSLLI 2011 Workshop on Projective Content*, 156–170. Columbus, Ohio: The Ohio State University.

Spector, Benjamin & Paul Egré. 2015. A uniform semantics for embedded interrogatives: *an* answer, not necessarily *the* answer. *Synthese* 192. 1729–1784. DOI: 10.1007/s11229-015-0722-4.

Stalnaker, Robert. 1978. Assertion. In *Syntax and Semantics*. P. Cole (ed.). Vol. 9. New York: Academic Press. 315–332.

Stalnaker, Robert. 2002. Assertion. In *Formal Semantics: the essential readings*. Paul Portner & Barbara H. Partee (eds.). Reprint of Stalnaker (1987). Oxford: Blackwell. 147–161.

Steedman, Mark J. 1990. Gapping as constituent coordination. *Linguistics and Philosophy* 13(2). 207–263. DOI: 10.1007/BF00630734.

Suñer, Margarita. 1993. About indirect questions and semi-questions. *Linguistics and Philosophy* 16(1). 45–77. DOI: 10.1007/BF00984722.

Suñer, Margarita. 1999. La subordinación sustantiva. La interrogación indirecta. In *Gramática Descriptiva de la Lengua Española*. Ignacio Bosque & Violeta Demonte (eds.). Madrid: Espasa Calpe. 2149–2195.

Taraldsen, Knut Tarald. 2000. Rethinking gapping and the order of constituents. In Øivin Andersen, Kjersti Fløttum & Torodd Kinn (eds.), *Menneske, språk og fellesskap*, 211–224. Oslo: Novus.

Tauscher, Elisabeth & Ernst-Georg Kirschbaum. 1970. *Grammatik der russischen Sprache*. 10th ed. Düsseldorf: Brücken-Verlag.

Terrell, Tracy. 1976. Assertion and Presupposition in Spanish Complements. In *Current studies in Romance linguistics*. Marta Luján (ed.). Washington, DC: Georgetown University Press. 221–245.

Terrell, Tracy & Joan Hooper. 1974. A Semantically Based Analysis of Mood in Spanish. *Hispania* 57(3). 484–494. http://www.jstor.org/stable/339187.

Thompson, Sandra A. & Anthony Mulac. 1991. The discourse conditions for the use of the complementizer that in conversational English. *Journal of Pragmatics* 15(3). 237–251. DOI: 10.1016/0378-2166(91)90012-M.

Tonhauser, Judith. 2016. Prosodic cues to presupposition projection. In Mary Moroney, Carol-Rose Little, Jacob Collard & Dan Burgdorf (eds.), *Proceedings of SALT 26*, 934–960. DOI: 10.3765/salt.v26i0.3788.

Tonhauser, Judith. 2020. Projection variability in Paraguayan Guaraní. *Natural Language and Linguistic Theory* 38(4). 1263–1302. DOI: 10.1007/s11049-019-09462-x.

Tonhauser, Judith, David I. Beaver & Judith Degen. 2018. How Projective is Projective Content? Gradience in Projectivity and At-issueness. *Journal of Semantics* 35(3). 495–542. DOI: 10.1093/jos/ffy007.

Toosarvandani, Maziar. 2016. Embedding the Antecedent in Gapping: Low Coordination and the Role of Parallelism. *Linguistic Inquiry* 47(2). 381–390. DOI: 10.1162/LING_a_00216.

Torrego, Esther. 1983. More effects of successive cyclic movement. *Linguistic Inquiry* 14(3). 561–565. http://www.jstor.org/stable/4178347.

van Craenenbroeck, Jeroen & Anikó Lipták. 2006. The crosslinguistic syntax of sluicing: evidence from hungarian relatives. *Syntax* 9(3). 248–274. DOI: 10.1111/j.1467-9612.2006.00091.x.

van Craenenbroeck, Jeroen & Anikó Lipták. 2008. On the interaction between verb movement and ellipsis: New evidence from Hungarian. In *Proceedings of the 26th West Coast Conference on Formal Linguistics*. Charles B. Chang & Hannah J. Haynie (eds.). Somerville, MA: Cascadilla. 138–146.

van Craenenbroeck, Jeroen & Anikó Lipták. 2013. What Sluicing Can Do, What It Can't, and in Which Language: On the Cross-Linguistic Syntax of Ellipsis. In Lisa Lai-Shen Cheng & Norbert Corver (eds.), *Diagnosing Syntax*. Oxford University Press. DOI: 10.1093/acprof:oso/9780199602490.003.0025.

van der Sandt, Rob A. 1992. Presupposition Projection as Anaphora Resolution. *Journal of Semantics* 9(4). 333–377. DOI: 10.1093/jos/9.4.333.

Vendler, Zeno. 1957. Verbs and times. *The Philosophical Review* 66(2). 143–160. http://www.jstor.org/stable/2182371.

Vicente, Luis. 2010. A Note on the Movement Analysis of Gapping. *Linguistic Inquiry* 41(3). 509–517. http://www.jstor.org/stable/40926392.

Villa-García, Julio. 2016. TP-ellipsis with a polarity particle in multiple-complementizer contexts in Spanish: on topical remnants and focal licensors. *Borealis* 5(2). 135–172. DOI: 10.7557/1.5.2.3781.

Villa-García, Julio. 2019. Recomplementation in English and Spanish: Delineating the CP space. *Glossa* 4(1). 56. DOI: 10.5334/gjgl.845.

Weir, Andrew. 2014. *Fragments and Clausal Ellipsis*. University of Massachusetts Amherst. (Doctoral dissertation). DOI: 10.7275/5823750.0.

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer. https://ggplot2.tidyverse.org.

Wilder, Chris. 1994. Coordination, ATB and Ellipsis. In *Minimalism and Kayne's asymmetry hypothesis*. Jan-Wouter Zwart (ed.). Groningen: Rijksuniversiteit. 291–329.

Wilder, Chris. 1996. V2-Effekte: Wortstellungen und Ellipsen. In *Deutsch - Typologisch*. Ewald Lang & Gisela Zifonun (eds.). Berlin, Boston: De Gruyter. 142–180. DOI: 10.1515/9783110622522-008.

Wilder, Chris. 1997. Some properties of ellipsis in coordination. In T. Alan Hall & Artemis Alexiadou (eds.), *Studies on universal grammar and typological variation*, 59–107. Amsterdam: John Benjamins.

Wolf, Lavi. 2014. *Degrees of assertion*. Ben Gurion University of the Negev. (Doctoral dissertation).

Wurmbrand, Susi. 2017. Stripping and Topless Complements. *Linguistic Inquiry* 48(2). 341–366. DOI: 10.1162/LING_a_00245.

Xue, Jingyang & Edgar Onea. 2011. Correlation between projective meaning and atissueness: An empirical study. In Gregory Kierstead (ed.), *Proceedings of ESSLLI 2011 Workshop on Projective Content*, 171–184. Columbus, Ohio: The Ohio State University.