Intervention effects in sentence processing

VILLATA, Sandra

Abstract

My dissertation investigates interference effects in the comprehension and production of long-distant dependencies in adults, integrating insights from theories on formal syntax and cognitive psychology granting a key role to similarity-based interference effects as a source of processing difficulty and/or ungrammaticality. My research mainly focuses on wh-islands’ acceptability and attraction effects in agreement, which can both be taken as windows on interference effects in sentence processing. My research suggests that: (i) only similarity on a narrow class of syntactic features has the potential to generate ungrammaticality, while similarity in terms of other features increases processing difficulty without changing the grammatical status of the sentence; (ii) attraction effects can be conceived as the result of similarity-based interference; (iii) encoding interference plays a key role in generating processing difficulties; (iv) self-organised sentence processing models offer a new promising way to implement interference effects in a rigorous mathematical framework.

Reference


DOI : 10.13097/archive-ouverte/unige:101927
URN : urn:nbn:ch:unige-1019279

Available at:
http://archive-ouverte.unige.ch/unige:101927

Disclaimer: layout of this document may differ from the published version.
INTERVENTION EFFECTS IN SENTENCE PROCESSING

Thèse de doctorat

Présentée à la Faculté des Lettres de l’Université de Genève pour l’obtention du titre de Docteur es Lettres, mention Linguistique

par

Sandra Villata

2017

Jury de thèse: Prof. Dr. Luigi Rizzi, directeur
Dr. Julie Franck, co-directeur
Prof. Dr. Ur Shlonsky, président du jury
Prof. Dr. Jon Sprouse, membre du jury
Dr. Patrick Sturt, membre du jury
Prof. Dr. Whitney Tabor, membre du jury
Preface

Chapter 2 reports research that was jointly conducted with Luigi Rizzi and Julie Franck. Chapters 3 and 4 report research conducted with Julie Franck. Chapter 5 reports research conducted with Julie Franck and Whit Tabor. Part of Chapter 2 was published with Luigi Rizzi and Julie Franck as co-authors in *Lingua*. Part of Chapter 3 was published with Julie Franck as co-author in *Rivista di Grammatica Generativa: Research in Generative Grammar (RGG)* and submitted as a co-authored journal article. Parts of the Chapter 4 were submitted together with Julie Franck as a co-authored journal article. Parts of the Chapter 5 were submitted together with Julie Franck and Whit Tabor as a co-authored journal article. Parts of all the chapters have been presented at various psycholinguistics and linguistics conferences from 2012 to 2017.
Acknowledgments

This thesis would not exist if not for the many people who lent me their support during these last six years. The gratitude, esteem and deep affection I feel for them are hardly expressible in words. These paragraphs represent a modest attempt to express these feelings.

First and foremost, I would like to thank my two advisors, Luigi Rizzi and Julie Franck, to whom I am incredibly indebted. I would never have written this thesis without their support and mentorship.

Luigi Rizzi supervised me with great attention while allowing me to pursue my project autonomously. This gave me the chance to find my own research path and develop my own ideas. I am extremely grateful for the chance I had to work with him and learn from him: he taught me almost everything I know about Linguistics and his ideas influenced my research deeply.

I would not be where I am today if Julie Franck had not believed that I would be able to pursue a PhD before I believed it myself. She has been an incredible source of inspiration for me in a variety of domains. I would not be the researcher and the woman I am now if our paths had not crossed. The affection and the profound esteem I feel for her cannot be expressed in a few lines. Hence, I will limit myself to this: Thank you, Julie!

I also owe a lot to Ur Shlonsky. He has always been extremely supportive during these years. I learned a lot from him, both from an academic and a human point of view. He has a truly refreshing and encouraging personality and I really enjoyed our friendly and intensely stimulating discussions throughout these years.

I also feel very lucky to have had the chance to meet Whit Tabor and collaborate with him. The months I spend at Uconn have been among the most beautiful, stimulating and cheerful moments of my PhD. His support was incredible and even after my visiting period at Uconn he never ceased to provide me with help and support in pursuing my goals in the Academia.

I would also like to thank all the members of the Psychology and Linguistic Departments at Uconn, who made my visiting period there possible and made it an unforgettable experience. I am also extremely thankful to Jon Sprouse for his valuable advice during these years, and Patrick Sturt for his support in the challenging moment that the transition between the PhD and what comes next represents. They are undoubtedly two of the researchers who influenced me the most.

I would also like to thank Akira Omaki, Brian McElree and Matt Wagers, with whom I had the incredible opportunity to collaborate at different moments of my PhD, and Roger Levy, Martin Corley, Shravan Vasishth, Andrew Nevins, Alessandro Treves, Julien Musolino, Adrian Staub, Klaus Oberauer, and Daniel Swingley, with whom I had the chance to exchange in several occasions. They are outstanding models of creativity and scientific rigor. What I learned from them cannot be quantified.

I feel lucky to have shared this experience with some great colleagues at the Laboratory of Psycholinguistics in Geneva. Special thanks go to Romy Lassotta, Daniele Panizza, and Audrey
Bürki, from whom I learned a lot and with whom I have shared many pleasant and funny moments. Thanks are also due to Giuliano Bocci for his support and friendship, and to all the members of the Department of Linguistics in Geneva. It is an amazing group, and I feel very lucky to have had the chance to share this experience with them. Thanks to Andres Posada and Alexandre Bourquin, my technical and E-prime gurus. I will miss the Genevan crew and this extremely pleasant work environment in the years to come!

Thanks to Giovanni Pastore, Marco Altoè and Davide Massidda: they were an enormous source of inspiration for statistical thinking and analyses with R at the summer school in Bertinoro in July 2016, undoubtedly the most incredible summer school I ever participated in. I would also like to thank Silvia Albertini for her friendship, but also for having initiated me to R in such an enjoyable way!

I am very indebted to Diego Marconi for initiating me into the study of language in the most rigorous and fascinating way I had ever seen. I think that if it was not for him, I would never have developed such a strong fascination for what I am doing. The memories of the philosophical years at the University of Turin are amongst my fondest ones.

I am very lucky to be surrounded by many extraordinary friends who have accompanied me in this amazing adventure, as well as in many others. Their friendship is priceless to me. Thank you Davide Pala, Dario Consoli, Ilaria Bada, Serena Cafagna, Salim Hirèche, Michele Ombrato, Fabrizio Calzavarini, Emiliano Boccardi, Giovanni Merlo, Cristian Marmo, Paolo Lorusso, Christos Christopoulos, and all of those I cannot mention for reasons of length, but I know they will recognize themselves in these lines.

This thesis is dedicated to my parents. Although they are still wondering how it is possible that I decided to pursue such a challenging path, how I can enjoy studying every single day and living far away from them and the Bel Paese, they have always let me follow my own path. This is the greatest gift of all. Mamma, Papà, grazie. Vi voglio bene.
# Table of Contents

**Chapter 1: Introduction**

1.1 The challenge of long-distance dependencies 1  
1.2 The syntactic viewpoint 5  
    1.2.1 Relativized Minimality 5  
    1.2.2 Featural Relativized Minimality 8  
    1.2.3 Featural Relativized Minimality: a new turn 11  
    1.2.4 Intervention and agreement 15  
1.3 The processing viewpoint 20  
    1.3.1 The cue-based memory model 20  
    1.3.2 Decay vs. Interference 24  
    1.3.3 Retrieval and Encoding interference 26  
1.4 Outline of the dissertation 31  
    1.4.1 Chapter 2: Intervention effects and Relativized Minimality 32  
    1.4.2 Chapter 3: A taxonomy of features 33  
    1.4.3 Chapter 4: Similarity-based interference in agreement features 34  
    1.4.4 Chapter 5: The mechanisms underlying interference effects 35  

**Chapter 2: Intervention effects and Relativized Minimality: New experimental evidence from graded judgments** 38  

2.1 Introduction 38  
2.2 Plan of the study 40  
2.3 Experiment 1: Graded judgments on wh-islands 45  
    2.3.1 Method 45  
    2.3.2 Results 47  
    2.3.3 Discussion 50  
2.4 Experiment 2: Binary judgments on wh-islands 50  
    2.4.1 Method 50
2.4.2 Results
2.4.3 Discussion

2.5 Experiment 3: D-linking or lexical restriction?
2.5.1 Method
2.5.2 Results
2.5.3 Effect of context: Comparison between Experiments 1 and 3
2.5.4 Discussion

2.6 General Discussion
2.6.1 Featural Relativized Minimality and weak islands
2.6.2 Cue-based memory models and similarity-based interference
2.6.3 Relativized Minimality and Superiority

2.7 Conclusions
2.8 Appendix

Chapter 3: Similarity effects in wh-islands acceptability: A taxonomy of features

3.1 Introduction
3.2 Present study
3.3 Experiment 1: The effect of animacy and lexical restriction in wh-islands
   3.3.1 Method
   3.3.2 Results
   3.3.3 Discussion
3.4 Experiment 2: The effect of reversibility and lexical restriction in wh-islands
   3.4.1 Method
   3.4.2 Results
   3.4.3 Discussion
3.5 Merge of Experiments 1 and 2
3.6 Experiment 3: The effect of animacy in wh-islands and that-clauses
   3.6.1 Method
   3.6.2 Results
   3.6.3 Discussion
3.7 Experiment 4: The effect of reversibility in wh-islands and that-clauses
   3.7.1 Method
Chapter 4: Similarity-based interference in agreement comprehension and production: Evidence from French object agreement

4.1 Introduction
4.1.2 Mismatch effects in production
4.1.3 Mismatch effects in comprehension
4.1.4 Accounting for the empirical variability: a proposal
4.2 Current study
4.3 Experiment 1: Number agreement in production
4.3.1 Method
4.3.2 Results
4.3.3 Discussion
4.4 Experiment 2: Gender agreement in production
4.4.1 Method
4.4.2 Results
4.4.3 Discussion
4.5 Experiment 3: Number agreement in comprehension 141
  4.5.1 Method 141
  4.5.2 Results 144
  4.5.3 Discussion 146
4.6 Experiment 4: Gender agreement in comprehension 149
  4.6.1 Method 149
  4.6.2 Results 151
  4.6.3 Discussion 153
4.7 General Discussion 156
  4.7.1 Attraction as errors of controller selection in sentence production 156
  4.7.2 Similarity-based interference from agreement features in sentence comprehension:
     Retrieval or encoding effects? 159
4.8 Conclusions 164

Chapter 5: The mechanisms underlying interference effects: Evidence from Italian and
English agreement 166
  5.1 Introduction 167
    5.1.1 Evidence of retrieval interference 169
    5.1.2 Evidence of encoding interference 172
  5.2 Experiment 1 179
    5.2.1 Method 179
    5.2.2 Results 181
    5.2.3 Discussion 184
  5.3 Experiment 2 185
    5.3.1 Method 186
    5.3.2 Results 188
    5.3.3 Discussion 191
  5.4 General Discussion 192
    5.4.1 Summary of the findings 192
    5.4.2 An encoding interference mechanism in ACT-R: Activation leveling 195
    5.4.3 A Self-Organized Sentence Processing (SOSP) account of encoding and retrieval
     interference effects 197
Chapter 6: Conclusions

6.1 Summary of the most relevant findings
   6.1.1 Intervention effects in wh-islands
   6.1.2 The mechanisms underlying the comprehension and production of grammatical long-distance dependencies
6.2 Further thoughts

References
List of Tables

Table 2.1. English translation of the French sentences in the 8 experimental conditions
Table 2.2. Summary of the interaction tests for the six models in Experiment 1
Table 2.3. Summary of the interaction tests for the six models in Experiment 2
Table 2.4. Example of context story and the corresponding questions for Experiment 3
Table 2.5. Summary of the interaction tests for the six models in Experiment 3
Table 2.6. Summary of the interaction tests for the four models in Experiments 1 and 3
Table 2.A.1. Mean ratings on a 7 points Likert scale in the 8 experimental conditions
Table 2.A.2. Fixed effects summary for Experiment 1 in the Intervention conditions
Table 2.A.3. Fixed effects summary for Experiment 1 in the No Intervention condition
Table 3.1. Summary of the fixed effects for Experiment 1
Table 3.2. Summary of the fixed effects for Experiment 2
Table 3.3. Summary of the fixed effects for the Experiment 1 and Experiment 2 merged together
Table 3.4. Summary of the fixed effects for the Experiment 3
Table 3.5. Summary of the fixed effects for the Experiment 4
Table 3.6. Summary of the fixed effects for the Experiment 5
Table 3.7. Summary of the fixed effects for Experiment 6
Table 3.A.1. Mean ratings on a 7 points Likert scale in the 4 experimental conditions in Experiment 1
Table 3.A.2. Summary of the fixed effects for Experiment 1
Table 3.A.3. Mean ratings on a 7 points Likert scale in the 4 experimental conditions in Experiment 2
Table 3.A.4. Summary of the fixed effects for Experiment 2
Table 3.A.5. Mean ratings on a 7 points Likert scale in the 4 experimental conditions in Experiment 6
Table 3.A.6 Summary of the fixed effects for Experiment 6
Table 4.1. Summary of the mismatch effects reported in production and comprehension studies as a function of Type of structure and Grammaticality. Complex structures differ from simple ones in that they involve movement. A detrimental effect of mismatch between the target and
the interfering element manifests in increased agreement errors in production and increased processing difficulty in comprehension, while a facilitatory effect manifests in reduced errors or processing difficulty times

**Table 4.2.** Example of item in the four experimental conditions of Experiment 1. The two past participles in the examples indicate the options among which participants were asked to make a choice

**Table 4.3.** Example of item in the eight experimental conditions of Experiment 2

**Table 4.4.** Example of item in the eight experimental conditions of Experiment 3

**Table 4.5.** Example of item in the eight experimental conditions of Experiment 4

**Table 5.1.** Example of item in the four experimental conditions of Experiment 1

**Table 5.2.** Mean accuracy percentages for comprehension questions by experimental condition in Experiment 1

**Table 5.3.** Example of item in the four experimental conditions of Experiment 2

**Table 5.4.** Mean accuracy scores of question responses in percentage by experimental condition in Experiment 2


List of Figures

**Figure 1.1.** Schematic illustrating the AGREE relation between the subject in its thematic position and AgrS (from Franck et al. 2006)

**Figure 1.2.** Schematic illustrating the Spec-Head relation between the subject in the AgrS’ specifier position and AgrS (from Franck et al. 2006)

**Figure 1.3.** Syntactic structure of a subject-verb agreement dependency in a structure involving object movement to the left periphery. $t_o$ refers to the trace of the moved object, $t_v$ refers to the trace of the verb (from Franck et al. 2010)

**Figure 1.4.** SAT function illustrating average d’ accuracy (i.e., hit rate minus false alarm rate) as a function of processing times in response to the 18 tones. Panel A shows the pattern expected for two conditions, X and Y, differing in their asymptotic accuracy, but with similar dynamics (same intercept and rate). Panel B shows the pattern expected for two conditions with identical asymptotic accuracy but different retrieval speed (intercept and rate)

**Figure 2.1.** Plot of the 8 individual structures on the 7-point Likert scale for Experiment 1

**Figure 2.2.** Plot of the 8 individual structures on the binary scale for Experiment 2

**Figure 2.3.** Plot of the 8 individual structures on the 7-points Likert scale for Experiment 3

**Figure 2.4.** Plot of the 8 individual structures on the 7-points Likert scale for Experiments 1 and 3. The first four structures on the left are conditions with No Intervention, while the four structures on the right are conditions with Intervention

**Figure 3.1.** Mean acceptability (z-scores) for the 4 conditions of Experiment 1. Error bars indicate standard errors

**Figure 3.2.** Mean acceptability in z-scores for the 4 conditions of Experiment 2. Error bars indicate standard errors

**Figure 3.3.** Proportion of animacy mismatch choices in that-clauses and wh-islands in Experiment 3. Error bars indicate standard errors. The horizontal dotted line indicates chance level

**Figure 3.4.** Proportion of selection of non-reversible sentences over reversible sentences in that-clauses and wh-islands in Experiment 4. Error bars indicate standard errors. The horizontal dotted line indicates chance level

**Figure 3.5.** Proportion of selection of that-clause sentences over wh-islands in animacy
mismatch, non reversible and animacy match, reversible conditions of Experiment 5. Error bars indicate standard errors. The horizontal dotted line indicates chance level.

**Figure 3.6.** Mean acceptability in z-scores for the 4 conditions of Experiment 6. Error bars indicate standard errors. The dotted line indicates fRM’s grammaticality cut off point of grammaticality, while the dashed line indicates the cut off point if only criterial features play a role in fRM, as I will propose below (see General Discussion).

**Figure 3.A.1.** Mean acceptability scores for the 4 conditions of Experiment 1. Error bars show standard errors.

**Figure 3.A.2.** Mean acceptability scores for the 4 conditions of Experiment 2. Error bars show standard errors.

**Figure 3.A.3.** Mean acceptability scores for the 4 conditions of Experiment 6. Error bars show standard errors.

**Figure 4.1.** Accuracy proportion in Experiment 1.

**Figure 4.2.** Distribution of responses times (in ms) in Experiment 1.

**Figure 4.3.** Accuracy proportions in Experiment 2.

**Figure 4.4.** Distribution of responses times (in ms) in Experiment 2.

**Figure 4.5.** Distribution of accuracy proportions to comprehension questions in Experiment 3.

**Figure 4.6.** Distribution of reading times (in ms) in the eight experimental conditions of the different regions of Experiment 3.

**Figure 4.7.** Accuracy proportion in the comprehension question of Experiment 4.

**Figure 4.8.** Distribution of reading times (in ms) in the eight experimental conditions for the different regions of Experiment 4.

**Fig. 5.1.** Distribution of reading times (in ms) in the four experimental conditions for the different regions of Experiment 1 (correct trials only).

**Fig. 5.2.** Distribution of reading times (in ms) in the four experimental conditions for the different regions of Experiment 2 for correct trials only (all subjects).
Chapter 1

Introduction

1.1 The challenge of long-distance dependencies

This dissertation is concerned with one of the most challenging phenomena the human parser is confronted with: long-distance dependencies between sentences’ constituents. Long-distance dependencies are syntactic/semantic relations between two constituents that are not adjacent in the sentence. Common examples include wh-questions, relative clauses, topicalizations and clefts. Long-distance dependencies (also known as unbounded or filler-gap dependencies) are the result of the unbounded nature of the human language that does not posit limits on the linear distance that such dependencies can cover. This is illustrated in the examples below, where the moved wh-object and its gap are in the same clause in (1a), but separated by one clause in (1b) and by two clauses in (1c). Of course, the greater the distance between the filler and the gap, the more cumbersome the resulting sentence, but this would not undermine its grammaticality.

(1)  a. **Which problem** did the linguist solve __?
    b. **Which problem** did you believe that the linguist solved __?
    c. **Which problem** did you believe that the linguist asked the psycholinguist to solve __?

The increasing difficulty in analyzing fully grammatical long-distance dependencies as a function of distance is usually attributed to the limitations of the memory system, which cannot keep track of an unlimited amount of material. The amount of information that can be actively maintained in memory (the *focus of attention*) is extremely limited, possibly to one single chunk (McElree and Dosher 1989; McElree 2006), while all the information that is outside the focus of attention needs to be retrieved from memory. Therefore, the human parser must be equipped, at a minimum, with a mechanism allowing it to retrieve previously encountered elements and integrate them with the incoming material. There has been an emergent consensus that the mechanism underlying this retrieval operation is a *cue-based, content-addressable* mechanism,
such that the target representation is accessed based on its content through retrieval cues triggered at the integration point (e.g., McElree 2000, 2006; Lewis and Vasishth 2005; McElree et al. 2006; Van Dyke and McElree 2006). Retrieval cues serve as pointers to the target representation allowing the parser to directly access the target based on the match between the retrieval cues and the content of the target, without the need to scan all the items in memory. Nevertheless, this extremely efficient retrieval mechanism has its drawbacks: in particular, it is highly sensitive to the interference from irrelevant representations also matching the retrieval cues, a phenomenon referred to as similarity-based interference (e.g., Anderson and Neely 1996; Van Dyke 2002, 2007; Van Dyke and Lewis 2003; McElree, Foraker and Dyer 2003; Lewis and Vasishth 2005; McElree 2006; Van Dyke and McElree 2006, 2011). This concept has proven to be extremely fruitful in accounting for the difficulty the parser encounters in processing a variety of long-distance dependencies, ranging from relative clauses to subject-verb agreement dependencies, ellipsis, cleft sentences, and so forth.

However, not all long-distance dependencies are of the kind illustrated above: some of them are not only hard to parse, but they also appear extremely degraded. A well-known example is represented by wh-islands, encapsulated environments that prevent the establishment of a long-distance dependency between an element located inside the island and an element outside of it (Ross 1967). In recent years, under the influence of Rizzi’s (1990) Relativized Minimality principle, wh-islands have started to be conceived in terms of intervention: in this approach, what induces a weak island effect is not the fact of having a particular construction in the island catalogue, the wh-island, but the fact of having an intervention configuration where the dependency between the filler and the gap is interrupted by the presence of an element in between which is similar in some relevant respects to the filler. To put it differently, Relativized Minimality reduces the concept of wh-island to the concept of intervention, which is a particular instance of the larger program of reducing properties of grammatical constructions to properties of finer computational ingredients.

In this dissertation, capitalizing on insights deriving from both formal syntax and cognitive theories of memory, I investigate how formal properties and structural operations interact with the memory operations that the parser deploys for the resolution of long-distance dependencies in real-time. This investigation raises several questions: which conditions generate sentence disruption as opposed to increased processing difficulty in long-distance dependencies? Are these two types of conditions distinct or is there a continuum between the two? Are these
conditions ultimately tied to the architecture of the human grammar or do they lie in the mechanism of the memory system? Are intervention effects and similarity-based interference the manifestation of a very same mechanism of mental computation or do they reflect separate constraints?

Relativized Minimality and the Cue-based memory model represent a privileged ground to pursue this investigation, given the commonalities that these theories share: they both grant a key role to the presence of intervening elements in accounting for the degradation and/or for the increased processing difficulty of sentences containing a long-distance dependency, and they both grant a key role to similarity, such that only elements that bear some similarity with the elements of the dependency have the potential to generate disruption and/or increased processing difficulties.

However, these two theories also differ in important respects. First, they pursue different explanatory aims: while Relativized Minimality, as a theory of grammar, aims to account for the degradation of sentences that are supposed to violate a rule of the grammar, the Cue-based memory model is concerned with the difficulty associated with the processing of fully grammatical sentences, even though this sharp division of labor has gradually become less and less clear-cut, a point to which I will return.

Second, the two theories are located at different levels of description. Using Marr’s (1982) terminology, Relativized Minimality is located at the computational level, as it is concerned with the formal properties and the structural operations that render a sentence well-formed, abstracting away from real-time operations. The Cue-based memory model is located at the algorithmic level, as it aims at describing the operations that the parser deploys in real-time in order to build the structure. Of course, a successful theory of language must deal with both levels of description plus a third one, the implementation level, which is concerned with the physical implementation of the linguistic computations in the human brain. The integration of these levels in a unified theory of language faces several challenges, the first being the linking hypothesis connecting these domains, which requires us to specify how the representations and the structure-building operations of the syntactic theory translate into parsing mechanisms and how these representations and mechanisms are realized in the brain (see Sprouse and Lau 2012; Poeppel and Embick 2005; Embick and Poeppel 2015). This dissertation is concerned with the first two levels of description and its goal is to show how the relationship between these levels
could be closer and more beneficial than it used to be. Ultimately, this dissertation bears on the long-standing question concerning the relationship between the grammar and the parser. I will defend the view that the functioning of the memory system does impose major constraints to sentence processing, modulating the ease with which elements of the sentence are encoded and retrieved from memory as the sentence structure is being built, thus accounting for processing difficulties. However, the Cue-based memory model does not account for conditions that lead the system to ultimately fail in resolving the long-distance dependency (e.g., wh-islands) as opposed to conditions that lead to increased processing difficulties without leading to ungrammaticality (e.g., object relative clauses), and it is unclear what characterization of similarity would allow this model to capture this distinction. I will suggest that in order to account for these observations, it is necessary to adopt a strong theoretical framework defining the structural conditions under which sentence disruption occurs and to provide a taxonomy of features that enables us to account for gradations in the strength of the disruption. I will argue that this framework is provided by Relativized Minimality. There are two types of empirical evidence that I offer in support of this view. In Chapters 2 and 3, I present experimental evidence showing that similarity in terms of some features, like the feature defining question operators (+Q), is particularly disruptive for sentence comprehension, unlike other features that only appear to mildly modulate sentence acceptability. I will argue that while sentence disruption should ultimately be understood in terms of constraints lying in the grammar, the mild effect due to similarity in terms of other features lies in constraints of the working of the memory system. In Chapters 4 and 5, I further explore the role of memory constraints in the processing of fully-grammatical long-distance dependencies, such as subject and object relatives, in which the similarity between the elements of the sentence is on agreement features. Importantly, even though the similarity effects observed in these conditions are tied to the functioning of the memory system, the formal constraints and the structural conditions under which these effects occur have to be understood within a precise syntactic framework.
1.2 The syntactic viewpoint

1.2.1 Relativized Minimality

Despite the unboundedness of long-distance dependencies, syntactic rules apply \textit{locally}, that is they apply on a restricted portion of the structure. One clear example of the locality of syntactic constraints is provided by the cyclic nature of long movement.

\[(2) \begin{array}{l}
\text{[CP Which problem\textsubscript{i} do [TP you believe [CP t_i that [TP the linguist solved \_\_]]]?]}
\end{array}\]

Despite the wh-object is far away from its canonical position (indicated by the underscore after the verb \textit{solved}), wh-movement applies in successive cycles of short movements: the wh-object first moves to the specifier of the most embedded clause, leaving an intermediate trace \((t_i)\), and then moves to its final landing site to the left periphery of the clause (Chomsky 1973). Stepwise movement is enforced by locality: it is because syntactic rules must apply locally that long movement has to occur in successive cycles. In this dissertation, I will focus on what has been referred to by Rizzi (2011) as \textit{intervention locality}, which occurs in the presence of an intervening element bearing certain properties making it a potential participant in the local relation between the extracted element and its trace.\footnote{In this dissertation, I will not discuss the concept of \textit{impenetrability locality}, which states that certain domains are impervious to rules and which partially corresponds to the concept of strong island (see Rizzi 2011 for a discussion).} In order to avoid any terminological confusion, it is worth noting that the term \textit{locality} is used to refer to two different levels of description: it is used both to refer to the vicinity of the elements in the sentence, and to the condition of application of syntactic rules. Hence, long-distance dependencies are non-local, in the sense that the filler and the gap are non-adjacent in the sentence, but syntactic rules apply locally, in the sense that they apply to the smallest structural domain in which a syntactic relation can be satisfied. To maximise clarity, I will use this term exclusively to refer to the locality of the syntactic rules.

The most prototypical case of intervention locality is provided by wh-islands. Consider the sentence in (3).
(3) *When do you wonder [who left __i ]?*

Since the seminal work of Rizzi (1990), wh-islands have been analysed in terms of intervention effects, to be captured by the Relativized Minimality principle. In its first version, the principle of Relativized Minimality was formulated as follows (see Rizzi 2013:172).

(4) In the configuration

\[
\ldots X \ldots Z \ldots Y
\]

\(X\) and \(Y\) cannot be connected by movement (or other local relation) if

(a) \(Z\) structurally intervenes between \(X\) and \(Y\);

(b) \(Z\) is of the same structural type as \(X\).

The first criterion (4a) of the principle defines the structural configuration of intervention: an element \(Z\) intervenes between \(X\) and \(Y\) when \(X\) c-commands \(Z\) and \(Z\) c-commands \(Y\). Therefore, an element \(Z\) that does not structurally intervene through c-command between \(X\) and \(Y\) does not undermine the well-formedness of the sentence. Compare sentences in (5) and (6) (examples by Rizzi 2013).

(5) When, did the uncertainty about who won dissolve __,?

(6) *When, do you wonder who left __,?

Both in (5) and (6) \textit{who} linearly precedes the trace of \textit{when}, but only in (6) it also c-commands it, therefore giving rise to sentence disruption. Thus, under Relativized Minimality (henceforth, RM), only an element hierarchically intervening in terms of c-command has the potential to block the establishment of a local relation. However, intervention by c-command is not sufficient to violate locality. Consider the sentence in (7):

(7) What, do you believe that John solved __,?

In (7), the subject \textit{John} intervenes by c-command between the moved element and its trace, but nonetheless the sentence is fully grammatical. The second criterion of the principle stated in (4b) defines an additional condition that the intervening element must meet in order to prevent the establishment of the local relation of the elements it separates: the intervening element must be of the same structural type as the extracted one. By \textit{structural type}, Rizzi (1990) referred to
the nature of the position occupied by the element, namely Head-position, A-position (i.e., argument positions in which theta-roles can be assigned, hence subject and object positions), and A'-position (i.e., non-argument positions, as operator positions which are occupied by elements that have already been theta-marked). For instance, an element targeting an A'-position cannot jump across an intervening element occupying an A'-position, while it can move across an intervening element occupying an A-position or a Head-position. Such a case is illustrated in (7) where John c-commands the trace of what, but nevertheless does not block the formation of the wh-dependency because it is not of the same structural type of the moved element what. The grammaticality of (7) is thus ensured by the fact that while what targets an A'-position, John does not. The case in which both criteria are met, and therefore the sentence is ungrammatical, is illustrated in (6): who intervenes by c-command and it is also of the same structural type of the extracted element (i.e., both are elements targeting an A'-position).

From now on, I will refer to locality violations that arise when both criteria (4a) and (4b) are met as intervention effects, while the term intervention will be used exclusively to refer to the structural configuration involving a c-commanding element intervening on a long-distance dependency (criterion (4a) only).

However, not all A'-interveners seem to block A'-chains. Consider the following contrast:

(8) *What, do you wonder who solved __?  
(9) Which problem, do you wonder who solved __?

The sentence in (9) is much less degraded than the sentence in (8) even though in both cases the intervening element and the moved one target an A'-position. This contrast suggests that not all A'-interveners block A'-chains. How does the theory account for this contrast? In its first formulation, RM accounted for it by assuming that wh-elements bearing a lexical restriction (e.g., which problem) can use non-local devices allowing them to escape the locality constraint and thus bind their traces (see Rizzi 1990 and Cinque 1990). Hence, the contrast above was accounted for by assuming two devices connecting wh-elements to their traces, one excluding the establishment of a local relation between two elements separated by an intervener, and one allowing lexically restricted extracted elements to escape locality. However, Starke (2001) put forth the idea that a single connecting device, modulated by the nature of the wh-elements at play, could capture this observation. In this view, similarity was defined in terms of the morphosyntactic featural specification of the elements and not in terms of the nature of the
position of movement (Rizzi 2001, 2004; Friedmann et al. 2009). I will refer to this set-theoretic approach as *Featural Relativized Minimality* (fRM).

### 1.2.2 Featural Relativized Minimality

The principle of Featural Relativized Minimality is formulated as follows.

(10) Given a configuration

\[ ... X \ldots Z \ldots Y \ldots \]

A local relation is disrupted between X and Y if

(a) Z structurally intervenes between X and Y;
(b) Z fully matches the specification in morphosyntactic features of X.

The catalogue of features that are relevant in this set-theoretical approach includes morphosyntactic features that have the potential to trigger movement (Rizzi 2001, 2004). This assumption is derived from a longstanding view that has been pursued in generative grammar according to which movement is always motivated by the need to satisfy some morphosyntactic requirement. In particular, it has been claimed that movement is motivated by feature checking (Chomsky 1995 and much subsequent work). Hence, the relevant features are [+Q], which defines wh-elements, [+R(el)], which defines the head of the relative clause, [+Top], which defines elements in a topic position, and [+Foc], which defines focalised elements. In addition, the [+N] feature associated with lexically restricted elements (e.g., *which NP*) was argued to also trigger movement, based on the observation that, in some languages, lexically restricted wh-elements target a different structural position than bare wh-elements: whereas the bare wh-object can remain *in situ*, the restricted wh-object has to move to a higher position in the tree. A straightforward case is offered by some Northern Italian Dialects (see Munaro 1999), where bare wh-elements end up in a clause-final position, while lexically restricted elements are clause-initial (see Friedmann et al. 2009: fn. 8). Munaro analyses the two cases as both involving leftward movement, but with a lower landing site for the bare wh-element, and a higher landing site for the lexically restricted wh-element. Further remnant movement of the IP to an intermediate position in the C-system determines the surface order in this analysis. Other languages offer other kinds of evidence. For example, in European Portuguese (Ambar 1988)
bare wh-elements require subject inversion, while lexically restricted elements do not. In Northern Norwegian dialects (Westergaard and Vangsnæs 2005), bare wh-elements do not trigger Verb Second, while lexically restricted wh-elements do. In Bavarian (Bayer and Brandner 2008), lexically restricted elements can co-occur with the complementizer dass, while bare elements cannot. Moreover, and perhaps even more relevantly, in a language permitting multiple wh-movement like Romanian, a lexically restricted non-subject wh-phrase can precede a subject bare wh-element (*Cu care candidat cine a votat? – For which candidate who voted*?), an ordering option that is typically excluded for bare non-subject wh-phrases, which would require the order subject wh - non-subject wh (*Cine cu cine a votat? – Who for whom voted*?). The exceptional ordering option with lexically restricted elements can be explained by assuming that they can target a higher position than bare elements in the left periphery (Alboiu 2002, Soare 2009; their analyses are in terms of D-linking, but they can be easily transposed in terms of lexical restriction).

Featural Relativized Minimality is therefore well-equipped to account for the contrast presented above, reported again for convenience:

(8) *What, do you wonder who solved ___ ?
(9) ? Which problem, do you wonder who solved ___ ?

The sentence in (8) is ungrammatical because it involves an intervening element (who) that carries the very same morphosyntactic featural specification as the moved element (what), since both bear one and the same feature, namely the [+Q] feature defining question operators. This configuration of feature identity is illustrated in (11a), where +A and +B represent morphosyntactic features. In contrast, (9) satisfies the principle of fRM, even if speakers still perceive a mild degradation, because the feature set of the intervener is included in the feature set of the moved element, as illustrated in (11b), and therefore the feature match is not complete, escaping locality effects. Finally, when the extracted element and the intervener do not have any feature in common, this leads to a configuration of disjunction, as showed in (11c).

(11) X Z Y
a. Identity: +A +A +A
b. Inclusion: +A,+B +A +A,+B
c. Disjunction: +A +B +A
For the sake of clarity, in (12) I provide an example for each of the set configurations reported in (11) specifying for each of them the featural specification of the relevant elements.

\[
\begin{array}{ccc}
X & Z & Y \\
\end{array}
\]

(12)  
\[\begin{array}{c}
a. \text{*What}_i \text{ do you wonder who solved } \_\_ \text{?} \\
\text{Identity} \\
[+Q] & [+Q] & [+Q] \\
\end{array}\]

\[
\begin{array}{ccc}
X & Z & Y \\
\end{array}
\]

(12)  
\[\begin{array}{c}
b. \text{?Which problem}_i \text{ do you wonder who solved } \_\_ \text{?} \\
\text{Inclusion} \\
[+Q,+N] & [+Q] & [+Q,+N] \\
\end{array}\]

\[
\begin{array}{ccc}
X & Z & Y \\
\end{array}
\]

(12)  
\[\begin{array}{c}
c. \text{What}_i \text{ do you think that he solved } \_\_ \text{?} \\
\text{Disjunction} \\
[+Q] & [+Pro] & [+Q] \\
\end{array}\]

In the set-theoretic approach put forward by fRM, structures are ruled out when the feature overlap between the intervener Z and the moved element X is full (12a), while fRM is satisfied when the feature overlap of the two elements is partial (12b) or null (12c). However, these three main types of set-theoretic relations do not exhaust all the possible set-theoretic configurations. Two additional relevant configurations are Inverse Inclusion, with the specification of Z properly including the specification of X, and Complex Identity, in which the identity of the featural specification between X and Z involves more than one feature, as illustrated in (13).³

\[
\begin{array}{ccc}
X & Z & Y \\
\end{array}
\]

(13)  
\[\begin{array}{c}
a. \text{What}_i \text{ do you wonder which student solved } \_\_ \text{?} \\
\text{Inverse Inclusion} \\
[+Q] & [+Q,+N] & [+Q] \\
\end{array}\]

\[
\begin{array}{ccc}
X & Z & Y \\
\end{array}
\]

(13)  
\[\begin{array}{c}
b. \text{Which problem}_i \text{ do you wonder which student solved } \_\_ \text{?} \\
\text{Complex Identity} \\
[+Q,+N] & [+Q,+N] & [+Q,+N] \\
\end{array}\]

Since fRM predicts a strong degradation when the featural specification of the intervener is identical to the featural specification of the extractee, Complex Identity (13b) should be as degraded as the identity configuration in (12a), to which I will refer from now on as Bare

² In Chapter 2, I will discuss how fRM accounts for the observation that Inclusion, although escaping locality effects, is nonetheless perceived as deviant to some extent.

³ Another relevant relation is Intersection, whose role is explored in Belletti et al. 2012, and Non-criterial inclusion, a discussion of which is provided in section 1.2.3.
Identity, since both wh-elements are bare (e.g., *what, who*), unlike Complex Identity where both elements are lexically restricted (i.e., *which NP*). However, the prediction that Complex Identity should pattern alike with Bare Identity, both being instances of feature identity, is already known to be invalidated by informally gathered judgments suggesting that Complex Identity is actually better than Bare Identity. Therefore, the principle of fRM should be amended in order to account for this observation, a point to which I will return in Chapter 2 (section 2.2). As far as Inverse Inclusion is concerned, fRM predicts this configuration to be as degraded as Bare Identity, since the featural specification of the intervener fully matches the featural specification of the extractee (the fact that it also bears an additional feature is irrelevant for the principle). The fRM principle thus has a directionality: what counts in the definition of locality is the full match between the featural specification of the intervener and the extractee, such that an extractee bearing one more feature as compared to the intervener satisfies the principle, while an intervener bearing one more feature as compared to the extractee does not.

### 1.2.3 Featural Relativized Minimality: a new turn

Featural Relativized Minimality was initially introduced to account for the degradation observed in wh-islands (and possibly other cases of weak islands) and the principle was conceived as a principle of grammar defining the boundaries of grammaticality: locality is violated, and the sentence is therefore ruled out by the grammar, when the feature overlap between the intervener and the extractee is full. However, in more recent years, fRM has been extended to also account for the difficulty observed in young children in the processing of certain A'-dependencies, such as object relative clauses and wh-questions, which are fully grammatical sentences.

There is a substantial amount of cross-linguistic evidence showing that young children struggle with the comprehension of object relative clauses, although they properly process subject relative clauses (see, among many others, Corrêa 1995; De Vincenzi, Arduino, Cicarelli and Job 1999; Arnon 2005, 2009; Friedmann, Belletti, and Rizzi 2009; Adani, van der Lely, Forgiarini, and Guasti 2010; Contemori and Belletti 2014; Bentea and Durrleman 2014). This difficulty mirrors the difficulty found in adults, who also show increased processing difficulties for object relatives as compared to subject relatives (see, among others, Frazier 1987;

Since Bever (1974), it is well known that the difficulty in the processing of complex sentences is significantly reduced by varying the syntactic type of the NPs (e.g., *The reporter the politician the commentator met trusts said the president won’t resign* is perceived as much harder than *The reporter everyone I met trusts said the president won’t resign*). Capitalizing on this insight, Gordon, Hendrick and Johnson (2001, 2004) conducted several self-paced reading experiments on subject and object relative clauses providing strong empirical evidence that the well-known disadvantage of object relatives as compared to subject relatives in adults (i.e., longer reading times at the relative verb and lower comprehension accuracy) was significantly decreased, if not eliminated, when the subject was either a pronoun or a proper noun and the object was a definite description (e.g., *The banker that John/you praised climbed the mountain*). These findings (also known as *NP*-type effects) were interpreted as supporting a model of working memory sensitive to similarity-based interference, where the higher the similarity between the elements of the sentence (e.g., the extracted object and the intervening subject), the stronger the interference effect is predicted to be, a point to which I will return in section 1.3.1.

In more recent years, and along partially independent lines, Friedmann, Belletti and Rizzi (2009) attested for these effects in young children as well. In particular, the authors found that Hebrew children aged 3;5-5;0 understood object relative clauses at chance when the moved object and the intervening subject were both lexically restricted (e.g., *Show me the elephant that the lion is spraying*), but they were significantly above chance when the moved object and the intervening subject differed in lexical restriction, as in free object relatives (e.g., *Show me who the lion is spraying*) and in headed object relatives with an impersonal arbitrary pro subject (e.g., *Show me the elephant that someone is spraying*). Similar findings were observed for object wh-questions, which were understood significantly above chance when the object was not lexically restricted (e.g., *Whom does the cat bite*?), but they were at chance when both the object and the subject were lexically restricted (e.g., *Which dog does the cat bite*?). The authors suggested that the failure of young children in processing object relatives may be accounted for in terms of intervention, in the framework of fRM. The authors started by the observation that object relatives instantiate a configuration of intervention: as illustrated in (14), the movement
of the object from its canonical position to the beginning of the sentence crosses the intervening 
c-commanding subject.

(14) Show me the elephant: that the lion is spraying _i_.

The second step consists in calculating the feature overlap between the intervener and the 
extractee. The sentence in (14) is a configuration of inclusion, since the featural specification 
of the intervening subject, i.e., [+N], is included in the featural specification of the extracted 
object, i.e., [+Rel], [+N]. The third step consists in comparing the featural specification of (14) 
with the featural specification of, for instance, headed object relatives with an impersonal 
arbitrary pro subject, as in (15) (but the same holds for free relatives): while (14) instantiates a 
configuration of inclusion, (15) instantiates a configuration of disjunction, since the intervening 
subject does not bear any of the features of the extracted object.

(15) Show me who, that the lion is spraying _i_.

From this observation, the authors conclude that, on the continuum from a maximum to a 
minimum of differentiation, disjunction is more distinct than inclusion, therefore accounting 
for the diminished processing difficulties in the former as compared to the latter. The proposal 
of Friedmann et al. (2009) thus consists in applying the machinery of fRM, which was initially 
conceived of as a principle of grammar, to account for children’s processing difficulties in fully 
grammatical sentences.

Two observations stand out. First, even though object relatives are harder to process than 
subject relatives for both children and adults, children often fail to process them, while adults, 
despite the processing difficulties associated with these structures, ultimately succeed in their 
interpretation. Second, object relatives instantiating a configuration of inclusion are not as 
degraded as wh-islands instantiating the same configuration set. I discuss these two 
observations in turn. As far as the first observation is concerned, namely the discrepancy 
between adults and children in the processing of object relatives, Friedmann et al. (2009) 
assume that even though the same principle of fRM is at play in children and adults, children 
would adhere to a stricter version of fRM, possibly because their lower working memory 
capacity imposes stronger constraints on the need for distinctiveness, such that only maximal 
distinctiveness (i.e., disjunction) can be successfully computed by children, while the adults’ 
system is satisfied with partial distinctiveness (i.e., inclusion). However, to the extent to which
the adults-children discrepancy is ultimately rooted in processing limitations that affect performance (Friedmann et al. 2009, p. 84; see also Goodluck 2010), there are good reasons to think that fRM equally applies to account for adults’ difficulties in object relatives. This represents a significant turn in the theory, since a principle that was initially conceived to account for the ungrammaticality of structures as wh-islands, it is now extended to also account for the processing difficulties associated with the computation of grammatical long-distance dependencies involving a configuration of intervention.

As far as the second observation is concerned, how does fRM capture the fact that feature overlap sometimes gives rise to strict ungrammaticality, while some other times it increases processing difficulties without undermining the grammaticality of the sentence? Rizzi (1997, 2004) suggested to distinguish between two families of features that would play a different role in generating intervention effects: criterial features, i.e., features able to trigger movement autonomously as [+Q], [+Rel], [+Top] and [+Foc], and non-criterial features, i.e., features contributing to the identification of the landing site of movement only when accompanied with a feature triggering movement, as [+N]. This proposal is further articulated by suggesting that the severity of the intervention effect is a function of features’s ability to trigger movement, such that an overlap on criterial features determines strong degradation, while an overlap on non-criterial features does not, even though it is associated with increased processing difficulties. To put it differently, Rizzi (1997, 2004) introduces a hierarchy of features articulated over three levels: (i) Type I (or criterial) features: features that trigger movement on their own, (ii) Type II (or non-criterial) features: features that trigger movement only if accompanied with a Type I feature, (iii) Type III features: features that do not trigger movement at all. Only an overlap on Type I features is supposed to generate a locality violation, while an overlap on Type II features does not violate locality, but it has the potential to increase processing difficulties, and an overlap of Type III features is supposed to have no effect at all.

Therefore, as a first approximation, we can distinguish between two cut-off points: an overlap in terms of Type I features determines ungrammaticality, while an overlap in terms of Type II features determines difficulties in the derivation of the sentence without undermining the grammaticality of it.4

---

4 For the time being, I leave aside the precise categorization of structures that do involve an overlap on a Type I feature, but that are nonetheless supposed to be rescued by the presence of
Equipped with this distinction, the difference in acceptability between object relatives and wh-islands is nicely captured by the principle. Object relatives with two full NPs (e.g., *The elephant that the lion is spraying is dancing*) instantiates a configuration of inclusion, similarly to wh-islands with lexically restricted extractee (e.g., *Which problem do you wonder who solved*?). However, while in the former case the feature overlap is on the non-criterial [+N] feature (i.e., +Rel,+N...+N), in the latter it is on the criterial [+Q] feature (i.e., +Q,+N...+Q). Since fRM assumes that the strength of the intervention effect is determined by the nature of the overlapping feature, the higher acceptability of object relatives as compared to wh-islands is captured by the principle. From now on, I will refer to the configuration of inclusion in the case of wh-islands as criterial inclusion (or Type I inclusion) and to the configuration of inclusion in the case of object relatives as non-criterial inclusion (or Type II inclusion).

1.2.4 Intervention and agreement

In more recent years, the concept of intervention has also been successfully applied to account for other performance data in agreement production (Franck, Lassi, Frauenfelder and Rizzi 2006; Franck, Soare, Frauenfelder and Rizzi 2010). Since the seminal work of Bock and Miller (1991), it has been shown that agreement errors occur when an element intervenes in the subject-verb agreement dependency, as in *The key to the cabinets are rusty* (example from Bock and Miller 1991). In these occasions, in a small but significant percentage of the cases, the intervening element “attracts” the agreement imposing its number on the verb, an effect known as number attraction. These effects have initially been conceived in terms of percolation (e.g., Vigliocco et al. 1995; Vigliocco and Nicol 1998; Franck et al. 2002): on this account, attraction errors arise due to the percolation of the number feature of the intervening noun up to the subject head, which is therefore incorrectly valued. An important assumption of the Feature percolation model is that only features inside the subject constituent can percolate through the syntactic tree and contaminate the subject root node. However, it soon became clear that attraction effects are not restricted to elements that modify the subject constituent, as they

a lexical restriction on the extractee, which is the case of wh-islands instantiating a configuration of inclusion, a point which I will return to in Chapter 3.
were also attested for adjuncts (e.g., *The grandmother, while talking to the girls, make the dough; Franck et al. 2004) and preverbal objects (e.g., *Il les prominent – He-S them-P walk-P, in Fayol et al. 1994; see also Hartsuiker et al. 2001; Hemforth and Konieczny 2004; Franck et al. 2006, 2010). Even more intriguingly, attraction effects were also attested in structures in which the attractor does not intervene between the subject and the verb in the linear structure, as it is the case for object relatives (e.g., Bock and Miller 1991; Franck et al. 2006, 2010; Staub 2009, 2010), object cleft sentences (Franck et al. 2006) and questions (e.g., *Are the helicopter for the flights safe?; Vigliocco and Nicol 1998). Bock and Miller (1991), who were the first to observe these effects in object relatives, did not assimilate them to the traditional attraction variety, but interpreted them as the result of a subject identification problem: on some proportion of trials, the complexity of the sentence may have rendered participants genuinely confused about which one of the two noun phrases is the subject of the verb (see also Staub 2009, 2010). However, Franck et al. (2006) assimilated these effects to the intervention variety by capitalizing on current syntactic analyses, which assume that the object passes through an intermediate position before landing in its final position to the left periphery of the clause. This intermediate object position has been claimed to be responsible for the intervention effect observed in subject-verb agreement dependencies.

According to the syntactic analysis of agreement (Pollock 1989, Belletti 1990, and much related works), there is a dedicated syntactic position, AgrS, which is involved in the subject-verb agreement relation, since it is in this position that the verb receives its inflection. Agreement is realized because AgrS enters into an AGREE relation with the subject in its thematic position (i.e., the specifier of the lexical verb). As a result of this operation, the person and number features of the subject are copied onto AgrS. The verb then moves to AgrS to receive its morphological specification of person and number.
Finally, the subject moves to the specifier of AgrS where, according to Franck et al. 2006, a second operation occurs, namely *agreement check*, which is computed on the local Spec-Head relationship between the subject in the specifier of AgrS and AgrS, as illustrated in Figure 1.2. This further operation is assumed because of independent findings showing that agreement tends to be more stable when the subject moves into the specifier of AgrS, i.e., pre-verbally, rather than when it remains in its thematic position, as noted by Guasti and Rizzi 2002 (e.g., *Many books are/*is on the table vs. *There are/is many books on the table*; example from Franck et al. 2006).

**Figure 1.1.** Schematic illustrating the AGREE relation between the subject in its thematic position and AgrS (from Franck et al. 2006).
**Figure 1.2.** Schematic illustrating the Spec-Head relation between the subject in the AgrS specifier position and AgrS (from Franck et al. 2006).

In the case of object movement, as it is the case for object relatives, current syntactic analyses assume that the object moves to an intermediate position, AgrO, which ensures gender and number object-past participle agreement in French (Kayne 1989, Chomsky 1995), an agreement dependency that will constitute the bulk of Chapter 4. As illustrated in Figure 1.3, this intermediate object position intervenes structurally in the AGREE relation between the verb in AgrS and the subject in its base position. The authors thus attributed the interference effects observed in object relatives and object cleft sentences to the intervention of the intermediate trace of the displaced object.

Franck, Soare, Frauenfelder and Rizzi (2010) provided compelling evidence in support of this analysis. The authors showed that when the object has not undertaken movement, as it is the case for patients in a complement clause structure (e.g., *John says to the patients that the medicine cures*), attraction errors are very rare, while they significantly increase when the object has moved to the front of the sentence, thus leaving an intermediate trace, as it is the case for patients in object relatives (e.g., *John speaks to the patients that the medicine cures*). Hence, two structures identical in their surface order exhibit very different attraction patterns, thus suggesting that what generates interference in the subject-verb agreement relation is the intermediate trace of the moved object and not the object in its final position. In this approach, attraction effects are thus conceived as yet another instance of intervention locality, where a local relation between two elements (e.g., the subject and the verb) fails in the presence of an intervening element (i.e., the intermediate trace of the object).
Figure 1.3. Syntactic structure of a subject-verb agreement dependency in a structure involving object movement to the left periphery, where $t_o$ refers to the trace of the moved object and $t_v$ refers to the trace of the verb (from Franck et al. 2010).

The question that arises next concerns whether similar interference effects are also detectable in the computation of agreement in sentence comprehension. Unlike results from agreement production, which are highly consistent in returning more attraction errors in mismatching conditions (i.e., when the agreement controller and the attractor mismatch in agreement features), results from comprehension are extremely mixed, ranging from penalty, through lack of effect, to facilitation. I will provide a thorough discussion of these effects and their implications in Chapter 4, while here I will restrict the discussion to the most relevant findings. Early studies in sentence comprehension shown that a prepositional phrase (PP) intervener mismatching in number the subject he ad (e.g., The key to the cabinets was rusty from many years of disuse) penalizes comprehension (i.e., longer reading times at the critical verb was, or at the immediately following region), thus mirroring results from production that also report more production errors in mismatch conditions (e.g., Nicol, Forster, and Veres 1997; Pearlmutter et al. 1999; Pearlmutter 2000). However, this effect is overturned when the sentence is ungrammatical (e.g., *The musicians who the reviewer praise so highly..., example from Wagers et al. 2009), in which case feature mismatch facilitates processing, with faster reading times at the verb region or at the region immediately after (e.g., Pearlmutter et al. 1999; Wagers, Lau, and Phillips 2009; Tucker, Idrissi and Almeida 2015; Lago, Shalom, Sigman,
Intriguingly, with the exception of a single study by Franck, Colonna and Rizzi (2015), no facilitatory mismatch effect is observed in grammatical structures, which is at odds with predictions from both the cue-based retrieval memory account and fRM, which predict a mismatch facilitatory effect to show up in grammatical sentences when the intervener mismatches the number of the target, due to reduced intervention / similarity-based interference effects. The widespread difficulty in observing such effect in the comprehension of grammatical agreement dependencies thus asks for an explanation (see Chapter 4). It also seems necessary to explain why agreement mismatch sometimes has a reverse effect in production and comprehension, which leads to the question of whether different mechanisms underlie comprehension and production. Franck et al. (2015) provide some preliminary answers to these questions. The authors suggested that the difficulty that previous studies in comprehension have shown in detecting a facilitatory mismatch effect in agreement dependencies may be due to methodological reasons: by using a mix of grammatical and ungrammatical sentences, these studies may have reduced the reliability of the number retrieval cue at the verb, therefore hiding the effect in grammatical conditions, a point I will return to in Chapter 4. The authors also suggested that the reverse mismatch effect sometimes observed in comprehension and production may be due to the process elicited by the task: performance in the comprehension task reflects the process of structure building (i.e., the resolution of an A’-dependency), which is hampered by an intervening matching element, while performance in production reflects the computation of agreement, which is hampered by an intervening mismatching element. However, this still does not explain why results from comprehension are not univocal, with some studies showing a facilitatory mismatch effect, some studies showing no effect and some other studies showing a detrimental mismatch effect. Moreover, since the study by Franck et al. (2015) is the only study showing a facilitatory effect of number mismatch in grammatical sentences, these results need to be replicated first. In addition, further theoretical elaboration is needed in order to provide a compelling account of the mechanisms at play in the comprehension and production of agreement dependencies accounting for why the mismatch effect is sometimes, but no always, reversed in comprehension as compared to production. I will be concerned with both issues (i.e., replication and theoretical elaboration) in Chapter 4.

1.3 The processing viewpoint
1.3.1 The cue-based memory model

Cognitive science has also approached long-distance dependency processing difficulties from the viewpoint of domain-general constraints from the memory system, rather than from domain-specific syntactic principles. However, although this approach assumes that overarching common principles, which are ultimately tied to the functioning of the memory system, subserve distinct cognitive domains, it also assumes that these principles manifest differently across cognitive domains, which captures the domain-specificity of such constraints, a point to which I will return shortly. Psycholinguistic evidence has shown that long-distance dependencies require the retrieval of the extracted element at some later point in the sentence, typically when the verb is encountered (see among others, Stowe 1986, Bever and McElree 1988, Osterhout and Swinney 1993). A very influential model that appears to be able to capture the retrieval mechanism underlying the processing of long-distance dependency is the Cue-based (or Content-addressable) retrieval model (e.g., McElree 2000; Lewis and Vasishth 2005; Van Dyke and McElree 2006; McElree 2006). According to this model, the retrieval operation is content-addressable, meaning that item information is retrieved based on cues that allow the parser to directly access the target element based on its content, rather than scanning all the elements in memory in sequence (e.g., McElree 2000, 2006; Van Dyke and McElree 2006). In this architecture, words are encoded in memory as vectors of features (e.g., Gillund and Shiffrin 1984; Nairne 1990; Oberauer and Kliegl 2006). Retrieval cues are triggered at the integration point (e.g., the verb) and form a subset of the features of the target (e.g., if the verb does not agree in gender, gender will not be a retrieval cue). The probability of retrieving a particular item in memory is defined as a function of the strength of the association between the retrieval cues and the features of the item.

The content-addressable retrieval model has been implemented by Lewis and Vasishth (2005) in a computational model of sentence processing realized in the ACT-R model (Anderson et al. 2004). In ACT-R the retrieval of an element is based on its activation. One of the main assumptions of the model is that each retrieval boosts the activation of the retrieved item. The activation of items in memory is therefore determined by: (i) the base activation level of the item, which is determined both by the past activations and re-activation of the item, and the time elapsed since the last retrieval, which lowers the item activation to a certain amount due
to decay, (ii) the strength of the association between the retrieval cue and the target element, that is the uniqueness with which the cue identifies the target, and (iii) a random noise component that allows retrieval not to be deterministic. On this view, the similarity-based interference from other elements in memory partially matching the retrieval cues reduces the activation of the target, thus accounting for increased processing difficulties in conditions of high interference, where multiple items in memory match the retrieval cues, as opposed to conditions of low interference, where only the target element matches the retrieval cues.  

The most compelling evidence in favor of a direct access mechanism comes from the speed-accuracy trade off (SAT) paradigm, which describes how accuracy varies as a function of retrieval time (Dosher 1979). In the SAT paradigm, words are automatically presented one after the other each 250 ms. Beginning 50 ms after the presentation of the last word, participants are asked to judge the grammaticality of the sentence (grammatical vs. ungrammatical) in response to 18 tones, one each 350 ms. This procedure provides three key parameters: (i) the asymptotic level of performance, which is a measure of the overall probability of retrieval, thus providing information relative to the strength of the memory representation, (ii) the intercept, i.e., the

---

5 I will not enter into the details of the ACT-R model, which fall beyond the scope of this thesis, but, for the sake of completeness, I point out that the Lewis and Vasishth (2005) model and the McElree (2000, 2006) model, despite both incorporating a cue-based, content-addressable retrieval mechanism, differ on a major point. Unlike the McElree’s model, the Lewis and Vasishth (2005) model does not assume a direct access mechanism, but rather incorporates assumptions of race models, according to which both retrieval probabilities and retrieval latencies are determined by the activation rate of elements in memory (e.g., Audley and Pike 1965; Vickers 1970). The main consequence of this assumption is that, while the McElree’s model assumes that retrieval speed is not affected by the amount of the interpolated material, only retrieval accuracy being affected by it, the Lewis and Vasishth’s model assumes that both retrieval accuracy and speed are affected by the interpolated material. This distinction is orthogonal to the issues at hand (I refer the reader to Nicenboim and Vasishth 2016 for a thorough discussion and an empirical evaluation of the two). In the remaining of this thesis, I will mostly refer to the McElree’s model that assumes a direct access to the relevant memory representation, while Chapter 5 will be mostly concerned with the Lewis and Vasishth ACT-R model.
moment in which performance starts to be different from chance (i.e., the moment in which participants start giving more correct than incorrect responses) and (iii) the rate at which accuracy grows from chance to the asymptote, as illustrated in Figure 1.4 (from McElree 2006). The advantage of the SAT procedure as compared to common reaction times paradigms is that it allows to distinguish between two independent factors that both influence reaction times: a process could be slower both because the retrieval speed is slower or because the quality of the target’s representation is degraded. While traditional measures of reading times, such as self-paced reading and eye-tracking, collapse these two sources of slowing down in reading times, the SAT paradigm allows us to tease apart the contribution of retrieval speed and retrieval accuracy.

**Figure 1.4.** SAT function illustrating average d’ accuracy (i.e., hit rate minus false alarm rate) as a function of processing times in response to the 18 tones. Panel A shows the pattern expected for two conditions, X and Y, differing in their asymptotic accuracy, but with similar dynamics (same intercept and rate). Panel B shows the pattern expected for two conditions with identical asymptotic accuracy but different retrieval speed (intercept and rate).

Studies making use of the SAT procedure combined with a probe recognition task (McElree 1996, 1998; McElree and Dosher 1989, 1993) show that: (1) retrieval dynamic is independent of set size (it takes the same time to retrieve a target in a set of 3 as in a set of 16 units), (2) retrieval dynamic is independent of the serial position of the item in the list (it takes the same
amount of time to retrieve a target presented at the end of the list as a unit presented at the beginning), and (3) the very last item memorized has a special status giving rise to faster dynamics, an element which has been referred to as the focus of the attention (McElree and Dosher 1989). These results provide strong support for direct access, since a search model would predict that the larger the list is, the longest the search, a prediction that is inconsistent with the finding that retrieval speed does not vary as a function of set size. Taken together, these observations point to a bipartite model of how recent events are represented: the last item is part of an attended state within focal attention and is accessed quickly, while all other items are part of a passive representational system, and are accessed more slowly, but with the same retrieval speed, through a content-addressable retrieval process (McElree and Dosher 1989; McElree 2006). This bipartite model contrasts with the tripartite architecture assumed in many current approaches, which assumes a focus of attention, a working memory system and a long-term memory system (e.g., Baddeley 1986; Baddeley and Hitch 1974; Shallice and Vallar 1990). One of the main results of studies employing the SAT procedure is that accessibility exhibits a sharply dichotomous pattern, therefore providing support for a bipartite memory model.

The SAT procedure has been extended to the study of memory for sentence comprehension, in which the probe recognition task was replaced by a sentence acceptability task. The first study investigating whether interpolated material between a long-distant constituent and its verb affected retrieval speed was conducted by McElree (2000). Using a SAT procedure, the author manipulated the distance between a moved object (the book) and its verb (admired) in a cleft construction by adding one or two subject relative clauses (e.g., *This was the book that the editor admired*). Participants were asked to judge the acceptability of the sentence. Grammatical sentences were intermixed with ungrammatical counterparts (e.g., *This was the book that the editor amused*). Results showed that even though the accuracy level varies as a function of the amount of interpolated material (i.e., the more the interpolated material, the lower the retrieval accuracy), retrieval speed was unaffected by it. The same results were observed in object clefts when the interpolated material was constituted by embedded complement clauses (e.g., *It was the scandal that the model believed that the celebrity relished*), but also in subject-verb dependencies when the interpolated material was constituted by a variety of different types of clauses (e.g., object relative clause, *The book that the editor admired ripped*; prepositional phrase and object relative clause, *The book from the
prestigious press that the editor admired ripped; object relative and subject relative clause, The book that the editor who quit the journal admired ripped; two object relative clauses, The book that the editor who the receptionist married admired ripped; see McElree et al. 2003), and in ellipsis (e.g., The editor admired the author’s writing, but [the critics] [everyone at the publishing house was shocked to hear that the critics] did not; Martin and McElree 2008, 2009). All in all, the studies above show that the amount of interpolated material does not affect the retrieval speed, but it does affect the probability of correct retrieval, such that the more the distractors, the lower the accuracy of retrieval.

1.3.2 Decay vs. Interference

The next question is which factor determines the asymptotic pattern. What does it make a target element less accessible in memory for retrieval, therefore reducing accuracy? Two obvious candidates are the passage of time, or decay, and interference. Decay refers to the fact that memory representations fade over time, thus becoming more and more difficult to retrieve as the time unfold, a notion that is central to the Baddeley Working Memory Model (Baddeley 2003; Baddeley and Hitch 1974). Interference refers to the fact that memory retrieval is impaired when interfering representations are present.

Evidence in favor of decay mainly comes from the studies from Brown 1958 and Peterson and Peterson 1959. These studies showed that the longer rehearsal strategies were suppressed (e.g., from 3 to 18 seconds), the lower the participant’s ability to recall a three-consonant trigram. However, subsequent studies called decay into question. Waugh and Norman (1965) varied the presentation rate of a list of digits that the participants had to recall later, such that the digits’ presentation was either slow or fast. Results showed that recall accuracy was not affected by the presentation rate, even if more time was passed between the item presentation and the recall test phase in the slow presentation condition as compared to the fast presentation condition (see Underwood and Keppel 1962 for similar results). Berman, Jonides and Lewis (2009) directly opposed predictions from decay and interference in a recent-probe detection task in which participants were asked to respond affirmatively if the probe was part of the current four-word set and negatively if it was not. To directly compare the effect of interference and decay, the authors designed three conditions in which the probe was not present in the last trial, but it was
present in a preceding set (i.e., negative probe): (1) the (negative) probe was present in the 2-back set; (2) the (negative) probe was present in the 1-back set which was followed by an inter-trial interval equated to the length of a single set (i.e., 10 seconds); (3) the (negative) probe was present in the 1-back set but the inter-trial interval was of only 1 second. In these three conditions participants had thus to respond negatively, since the probe was not in the current trial. The authors reasoned that the comparison between conditions (1) and (2) provided a test for interference, while the comparison between conditions (2) and (3) provided a test for decay. On the one hand, decay predicts that the more time has passed between the negative probe and recall, the lower the interference, because the negative interfering probe would have decayed due to the passage of time. Decay thus predicts no difference between conditions (1) and (2), which should be in turn faster than condition (3). On the other hand, interference predicts that the greater the number of elements between the negative interfering probe and recall, the lower the interference from the negative probe, because its activation would have been reduced by the presence of other interfering elements. Interference thus predicts no difference between conditions (2) and (3), which should be in turn slower than condition (1). Results attested that participants were faster in responding negatively in condition (1) as compared to conditions (2) and (3), but no difference was attested between these two latter conditions, therefore providing strong support for the interference account.

Evidence for interference effects independent of decay have been attested in a variety of extra-linguistic domains (see Lewis 1996 for a discussion). For instance, interference effects are attested in visual memory tasks: distractors that look perceptually similar to the target generate more interference than dissimilar ones (Magnussen, Greenslee, Asplund, and Dyrnes 1991). Similar evidence has also been gathered for odor memory: recognition performance for odors was poorest when a distracting odor were presented in the retention interval, prior to the recognition phase (Walk and Johns 1984). Hence, interference effects during retrieval are not unique to language, but they are rather tied to domain-general mechanism lying in the functioning of the memory system. However, interference manifests differently depending on the cognitive domain under investigation: in visual memory, for instance, it is the perceptual similarity that generates interference, while, as I will discuss in the next section, perceptual similarity does not appear to modulate interference in language comprehension. Rather, it is the similarity in linguistic properties that generates interference in sentence comprehension. Hence, interference effects due to similarity, despite being tied to a common overarching principle that
lies in the functioning of the memory system, manifest differently across cognitive domains due the specificities of the domain under investigation.

1.3.3 Retrieval and Encoding interference

Unlike decay, which is concerned with the quantity of information, interference is concerned with the quality of information: it is the distinctiveness of the retrieval cues that determines the probability of successful retrieval. Indeed, the probability of retrieval is determined by the degree of match between the retrieval cues and the target element, but also by the resonance between the retrieval cues with other elements in memory: the more elements in memory match the retrieval cues, the less accurate the retrieval process is expected to be, since the retrieval cues would match multiple items in memory (e.g., McElree 2000; McElree et al. 2003; Van Dyke and McElree 2006; McElree 2006; Van Dyke 2007; Van Dyke and McElree 2011). Direct evidence for similarity-based interference at retrieval comes from a study from Van Dyke and McElree (2006). In a memory load experiment, the authors manipulated the retrieval cues at the verb in object-cleft sentences (e.g., *It was the boat that the guy who lived by the sea sailed/fixed in two sunny days – Memory load: table, sink, truck*), which could either match or not the semantic content of the memory load (i.e., each element of the memory load is ‘fixable’, but none of them is ‘sailable’). The authors observed a slow-down in reading times at the verb when the semantic cue at the verb matched both the semantic feature of the target and those of the elements in the memory load (the fixable condition) as compared to when they only matched the features of the target (the sailable condition), which attests to similarity-based interference at retrieval. Other studies showed that in sentences in which the subject (*the lady*) and the verb (*moaned*) are separated by a relative clause, a constituent inside the relative clause (*the seat*) engenders significantly more interference when it also occupies a subject position, similar to the target of retrieval (e.g., *The pilot remembered that the lady who said that the seat was smelly yesterday afternoon moaned about a refund*), than when it occupies a prepositional object position (e.g., *The pilot remembered that the lady who was sitting in the smelly seat yesterday afternoon moaned about a refund*; Van Dyke and Lewis 2003, Van Dyke 2007). Here, it is similarity in terms of syntactic roles that appears to penalize the system, which suggests that structural properties (i.e., being in a subject position) also drive the retrieval process. Van Dyke 2007 also found, in the same structures, that when the intervening element was a semantically
plausible subject for the verb (in virtue of its animacy), sentence comprehension was significantly slower and less accurate, even if the intervening element did not have the same syntactic status (i.e., subject) as the target, suggesting that semantic similarity played a role independently of syntactic similarity. Nevertheless, recent evidence from Van Dyke and McElree (2011) suggests that the same semantic variable fails to influence retrieval if the intervening element occupies a direct object position, suggesting that although both syntactic and semantic cues play a role in sentence comprehension, syntactic cues are weighted more strongly in the retrieval process, and possibly play a gating function so that semantic interference effects can only be engendered by distractors with matching syntactic properties.

Further evidence has shown that both linearly and structurally intervening elements affect retrieval accuracy. McElree, Foraker, and Dyer (2003) manipulated the distance between the extracted element and its retrieval site in cleft sentences, by adding either a linearly intervening embedded relative clause (e.g., *This was the book that the editor who the receptionist married admired*) or a hierarchically intervening complement clause (e.g., *It was the scandal that the model believed that the celebrity relished*). Using the SAT procedure, the authors replicated the well-known decrease in retrieval accuracy as a function of the amount of interpolated material.

Interestingly, the same effect on retrieval accuracy was found for the linear and the hierarchical interpolated structures, suggesting that retrieval accuracy is not sensitive to the syntactic nature of this material. In line with this, studies have shown similar interference from elements coming before the target (i.e., proactive interference; e.g., Gordon et al. 2002, Van Dyke and McElree 2006) and from elements coming after the target (i.e., retroactive interference; see among others McElree 2000, McElree et al. 2003, Van Dyke and Lewis 2003), even though there is some empirical evidence suggesting that retroactive interference may be stronger than proactive interference (Martin and McElree 2009; Van Dyke and McElree 2011).

Another set of studies has shown that similarity-based interference effects also arise independently of retrieval. In a series of self-paced reading experiments, Gordon et al. (2001, 2004) showed that the well-known advantage found for subject relatives as compared to object relatives is reduced (or even eliminated) when the extracted object and the intervening subject are syntactically dissimilar, for example, when the object is a definite description and the subject is either a pronoun or a proper name (see Arnon 2010 for similar evidence in children).

Since the verb does not subcategorize for the syntactic type of its arguments (i.e., definite description vs. proper noun vs. pronoun), the syntactic type of the arguments cannot be a
retrieval cue. Therefore, these effects have been attributed to similarity-based interference taking place at encoding, that is, during memory storage and maintaining. It has been claimed that encoding interference may be the result of a mechanism of feature overwriting: under the hypothesis that features are represented as unique units, which cannot therefore belong to more than one item at the time, elements sharing a feature are assumed to enter in a competition for that feature and the element losing the competition will lose the feature, as well as a certain amount of its activation, therefore resulting in a degraded representation (e.g., Nairne 1990, Oberauer and Kliegl 2006). The assumption that multiple elements cannot share a feature is based on neural network models that assume that features are represented by oscillating units (e.g. Raffone and Wolters 2001). Features that belong to the same item oscillate in synchrony, while features that belong to different items oscillate out of synchrony. Since one same feature cannot oscillate in two different phases at the same time, it follows that one feature cannot belong to one than more item at the same time. Hence, whenever two or more items in memory share a feature, only one item will ultimately have it.

Crucially, interference taking place at encoding can also impact retrieval: when the activation and the quality of the memory representation decrease, possibly due to a mechanism such as feature overwriting, retrieval is more difficult because less active items in memory are more difficult to be accessed, a point that I will carefully discuss in Chapter 5 (see also Jäger, Benz, Roeser, Dillon and Vasishth 2015). Evidence going in this direction is provided by a recent study by Hofmeister and Vasishth (2014). In two self-paced reading experiments, the authors showed that semantically and syntactically more complex targets are retrieved faster than simpler ones at the target verb, even though they take more time to be encoded (i.e., longer reading times are associate with complex than simple elements, even when word length is controlled for in statistical analyses) (e.g., The congressman interrogated the general/ the victorious four-star general who a lawyer for the White House advised to not comment on the prisoners). However, since target complexity is not a retrieval cue, the effect has been attributed to similarity-based interference arising at encoding, even though the effect manifests at retrieval. In particular, more specific (i.e., complex) elements are likely to be more dissimilar one from the other, both semantically and syntactically, therefore reducing interference at encoding and facilitating retrieval. The authors also shown that not any feature enters in the computation of similarity. In a second self-paced reading experiment, the authors manipulated the color of the target which could either match or mismatch the surrounding text. If any feature
can trigger similarity-based interference at retrieval, one would expect lower interference in the color mismatch condition than in the color match condition, which would manifest in faster retrieval in the former condition. Results show no effect of color at the retrieval region, therefore suggesting that the catalogue of features relevant for retrieval is limited to features that are linguistically relevant (i.e., features that are likely to drive the parser during retrieval, such as syntactic and semantic features). However, an effect of text color congruency was reported during the encoding of the second noun phrase, but the effect goes in the opposite direction from what is expected under the hypothesis that feature overwriting can also affect superficial non-linguistic dimensions, since congruent conditions (i.e., conditions in which the target element has the same color as the surrounding text) are read faster than incongruent conditions (i.e., conditions in which the target element has a different color as the surrounding text), a result that the authors do not address.

Similar evidence comes from Kush, Johns and Van Dyke 2015, who showed that phonological similarity (i.e., the fact that the target rhymes with other elements in the sentence) does not have a role in the resolution of long-distance dependencies, i.e., during the retrieval of the long-distance element, even though a detrimental effect of phonological similarity was observed at encoding, i.e., during memory storage, with longer reading times at the second noun phrase region when it rhymes with the first noun phrase. Results from Kush et al. are at odds with findings from Acheson and MacDonald (2011), who found evidence for phonological interference effects both at the second noun phrase region and at the three following regions, which included the relative verb. However, Kush et al. observed that the sentences used by Acheson and MacDonald were characterized by overlapping phonological words throughout the sentence (e.g., *The baker that the banker sought bought the house*, where both the object and the subject of the relative clause and the embedded and the main verbs phonologically overlap), therefore preventing the possibility to establish whether the slow down observed at the verb occurred during the retrieval of the verb’s arguments, or during the encoding of phonologically similar elements. Kush et al. thus suggested that, based on their results that allow us to disentangle encoding and retrieval interference, also findings from Acheson and MacDonald should be interpreted in terms encoding interference, contra Acheson and MacDonald who proposed an interpretation in terms of retrieval interference.

Even though the studies reported above all provided evidence for encoding interference, the effect is sometimes detected at the encoding region (i.e., when the interfering noun phrase is
encountered), while some other times it is detected at the retrieval region (i.e., at the verb). In particular, Hofmeister and Vasishth observed an effect of target’s complexity at the retrieval region\textsuperscript{6}, but the effect of color was only detected at the encoding region, similarly to the phonological effect observed by Kush et al., which was also only observed at the encoding region.\textsuperscript{7} One possibility to address this variability in the results is that semantic information, such as that carried by complex noun phrases, may be longer-lasting than phonological information, which would explain why only the former manifests in later regions of the sentence and, in particular, at the retrieval region. This hypothesis is supported by the observation that semantic information, but not phonological information, is involved in sentence recall (Potter and Lombardi 1990). On this view, semantic-conceptual information is longer-lasting, possibly because this information is directly relevant to efficiently built the structure, while more superficial information, being irrelevant to that purpose, would tend to decay first. I will return to these observations in greater details in Chapter 5. For the time being, it is sufficient to stress that \textit{not any} feature can generate similarity-based interference effects at retrieval when shared both by the target and the intervener, but only those features that are relevant for structure building operations during language comprehension have this potential. Moreover, I also discussed that an interference effect observed at the retrieval region can be due both to encoding and retrieval interference, since the mechanism supposed to underlie encoding interference (i.e., feature overwriting), by lowering the activation level of the element that has lost the competition, also negatively impacts retrieval by decreasing retrieval probabilities, an aspect to which I will return in Chapter 5.

The question that arises next is whether linguistic features are all weighted the same in generating similarity-based interference effects or not. The cue-based memory model is not equipped with a system allowing it to attribute a different weight to different linguistic features.

\textsuperscript{6} As mentioned above, Hofmeister and Vasishth (2014) also observed an effect at encoding such that complex noun phrases were associated with longer reading times than simple noun phrases. However, this effect was driven by the fact that a complex noun phrase bears additional syntactic and semantic features, which take time to be encoded thus leading to a slow down in reading times (see also Hofmeister et al. 2007, 2013 for similar findings).

\textsuperscript{7} I put aside results by Acheson and MacDonald (2011) for which no clear conclusion can be drawn concerning the locus of the interference effect for the reasons detailed above.
It is unclear, for instance, how this model can account for why feature similarity causes ungrammaticality in wh-islands, while it only increases processing difficulties in object relative clauses. It is not impossible, in principle, that a hierarchy of features could be motivated exclusively on the basis of cognitive-based principles, but to date it is still unclear how this goal could be attained. I will return on this point in Chapter 3.

Summarizing, the Cue-based retrieval model accounts for the difficulty in processing long-distance dependencies by assuming that the retrieval operation at play to retrieve and integrate that long-distant constituent is sensitive to similarity-based interference. Interference is caused by elements that are similar to the target either in their syntactic or semantic features and can have an impact at encoding, at retrieval or both.

1.4 Outline of the dissertation

The dissertation consists of two main parts. Chapters 2 and 3 are concerned with intervention effects in structures commonly assumed to be ungrammatical, i.e., wh-islands, and provide a systematic investigation of how different degrees of feature overlap modulate the acceptability of wh-islands by integrating insights from both Featural Relativized Minimality and the Cue-based retrieval model. The main conclusion of these chapters is that two families of features should be distinguished: those responsible for setting the boundaries of grammaticality, such that an overlap in these features causes sentence ungrammaticality, and those that operate at the level of processing, therefore modulating the ease with which elements are encoded and/or retrieved from memory, without undermining the grammaticality of the sentence. Chapters 4 and 5 build upon this conclusion and investigate the memory processes that are affected by similarity in features belonging to the latter family of features (i.e., those features that I have argued to operate at the level of processing). Results from the experiments presented in both chapters provide strong evidence for similarity-based interference arising at encoding, possibly as a result of a mechanism of feature overwriting. These results represent a challenge for the cue-based retrieval memory model (e.g., McElree 2000, 2006; Lewis and Vasishth 2005) that, in its current formulation, does not include a mechanism able to generate interference at encoding. To address this issue, I first provide a discussion of how these effects may be reconciled with a cue-based memory architecture by introducing a new mechanism in the
model, activation leveling, and then I discuss how a self-organized sentence processing model can account for interference effects (Chapter 5).

1.4.1 Chapter 2: Intervention effects and Relativized Minimality

In three acceptability judgment studies investigating intervention effects in wh-islands, the feature overlap of the intervener and the target was systematically manipulated, giving rise to four configurations: Bare Identity, involving a full overlap of features, Inverse Inclusion, in which the featural specification of the intervener is richer than the featural configuration of the target, Inclusion, in which the featural specification of the target is richer than the featural specification of the intervener, and Complex Identity, involving a full overlap on more than one feature. Featural Relativized Minimality predicts full overlap of features to be particularly disruptive, while a partial overlap of features should lead to only mild deviance. I first summarize the basic facts that motivated the FRM account and I then discuss in detail FRM’s predictions for wh-islands configurations. An additional hypothesis is adopted for Complex Identity, which is argued to reduce to a configuration of Inclusion under the assumption that the restricted wh-extractee targets a complex head while the intervening one targets a simple head, the parser acting opportunistically in its attempt to maximize well-formedness. I also contrast predictions deriving from an approach that identifies the crucial factor distinguishing bare and complex wh-elements in a purely formal property, i.e., lexical restriction, with an approach that advocates an interpretative property, i.e., Discourse-linking. Results from the three experiments are highly consistent: predictions from Featural Relativized Minimality are globally borne out, systematically returning Inclusion and Complex Identity to be rated higher than both Bare Identity and Inverse Inclusion, which were found to be on a par. However, Complex Identity was systematically rated higher than Inclusion, a result that is unexpected under the hypothesis that Complex Identity reduces to a configuration of Inclusion. An interpretation in terms of processing is proposed, where more complex elements increased retrieval probabilities because of their richer semantic features that reduce similarity between the target and the intervener, therefore increasing acceptability judgments. Finally, results support the hypothesis that the ameliorative role ensured by complex extractees should be seek in lexical restriction rather than in D-linking.
The experimental and theoretical work reported in this chapter is the result of several years of fruitful collaboration with Julie Franck and Luigi Rizzi. Experiments 1-3 are reported in Villata, Rizzi, and Franck (2016).

1.4.2 Chapter 3: A taxonomy of features

Chapter 3 takes up a discussion of the implication of results presented in Chapter 2. I start with the observation that Complex Identity was found to be rated higher than Inclusion, an observation that has been accounted for in terms of increased distinctiveness between the target and the intervener due to the semantic richness of restricted elements. However, this solution results in maintaining that while the ameliorative role of lexical restriction in the case of Inclusion stems from a fRM-type effect, its role in the case of Complex Identity is external to fRM. Although logically possible, this solution does not appear to be very parsimonious, as it multiplies the explanations to account for the increase in acceptability judgments due to lexical restriction. In Chapter 3, I explore an alternative solution which consists in adopting a different taxonomy of features, where features triggering movement (i.e., criterial Type I features) play a role in defining the boundary of grammaticality, while all other linguistic features (i.e., Type II and Type III features) play a role in memory retrieval and encoding. According to this view, Featural Relativized Minimality is a theory defining the boundaries of grammaticality, while the fine-grained modulation observed among the wh-island configurations is assumed to be the result of extra-grammatical factors which are tied to the functioning of the memory system. In other words, this approach restricts the set of features relevant in the calculation of the locality effect to those referred to as criterial by Rizzi (2004), and it considers lexical restriction in the same way as other features not triggering movement.

This approach is tested in six acceptability judgment studies in French, in which the similarity between the extractee and the intervener was manipulated in wh-island structures and in minimally different grammatical structures (extraction out of a that-clause). Four features were manipulated: a Type I feature, i.e., the +Q feature defining question operators, a Type II feature, i.e., lexical restriction (+N), and two Types III features, i.e., animacy and reversibility of thematic roles. Experiments 1 and 3 test, by way of an acceptability judgment procedure and a two-alternative forced choice task, the role of animacy in modulating the acceptability rates of
both wh-islands and that-clauses. Experiments 2 and 4 test the role of thematic role reversibility in both wh-islands and that-clauses, with the same procedures of Experiments 1 and 3. Experiment 5 tests the effect of an overlap in the criterial +Q feature in a two-alternative forced choice task, while Experiment 6 provides a direct comparison of the effect size of an overlap in the +Q feature and an overlap in the +N feature.

Results show that lexical restriction, animacy and reversibility of thematic roles all mildly but significantly contribute in modulating acceptability rates. In particular, their effect is much reduced as compared to the effect of the criterial +Q feature, and it is found both in island and non-island configurations. I further discuss other findings in the literature suggesting that these features also affect structures without intervention, such as subject relatives, as well as grammatical structures with intervention.

The experimental and theoretical work reported in this chapter is the fruit of the close collaboration with Julie Franck. Experiments 1-6 are reported in Villata and Franck (submitted).

1.4.3 Chapter 4: Similarity-based interference in agreement features

Chapter 4 builds upon the claim developed in Chapter 3 that similarity effects due to similarity in Type II and III features are tied to the functioning of the memory system, and thus operate at the level of processing by modulating the ease with which sentential elements are encoded or retrieved from memory during structure building. The first aim of Chapter 4 is to extend this investigation by testing the role of agreement features in modulating similarity-based interference in agreement dependencies. Agreement features are expected to belong to the family of features that have the potential to increase sentence processing difficulties when shared both by target and the intervener without undermining the grammaticality of the sentence. However, this expectation is challenged by several studies reporting no effect of similarity in agreement features in modulating the comprehension of grammatical sentences such as object relatives (e.g., Wagers et al. 2009, Dillon et al 2013, Lago et al. 2015), an observation that is at odds with standard reports of reduced similarity-based interference in conditions of feature mismatch. In this Chapter, I further investigate the role of agreement features in modulating the comprehension of fully-grammatical object relative clauses, paving the way for a reconciliation of apparently contradictory findings. With this aim, I investigate
sentence comprehension in object-past participle agreement in French with a self-paced reading procedure: Experiment 3 manipulates number mismatch, while Experiment 4 manipulates gender mismatch. Results show that number and gender match significantly decrease comprehension accuracy of object relatives, in line with well-attested similarity-based interference effects observed in the processing of long-distance dependencies. Intriguingly, however, a comparable detrimental effect of agreement feature match is also observed for subject relatives, which were used as control. Since these structures do not contain a long-distance dependency, the observed effect cannot lie in the mechanisms of memory retrieval. I suggest that this effect is likely to originate at encoding, possibly through a mechanism of feature overwriting. Moreover, since the size of the effect was comparable for subject and object relatives, I suggest to interpret the mismatch effect observed in object relatives as also stemming from encoding interference.

The second aim of Chapter 4 is to provide a unified account of findings from agreement production and findings from agreement comprehension. Even though language comprehension and production have traditionally been treated as independent systems (e.g., MacDonald, Bock, and Kelly 1993; Dell, Burger, and Svec 1997; Levelt et al. 1999), unified models assuming that they share similar mechanisms have started to develop (e.g., Pickering and Garrod 2013, Dell and Chang 2014). On this view, one may expect interference effects to manifest similarly in comprehension and production. Agreement dependencies represent a privileged ground to address this question, since interference effects in agreement dependencies are observed both in comprehension (e.g., Pearlmutter et al. 2009, Wagers et al. 2009, Franck et al. 2015) and in production (e.g., Bock and Miller 1991), even though their effect is sometimes reversed, an observation that I also discuss in this chapter. Experiments 1 and 2 investigate agreement production with a force-choice response time paradigm with a Rapid Serial Visual Presentation (RSVP) procedure, in which sentences are rapidly and automatically presented in a word-by-word fashion and participants are asked to select the verb form that correctly completes the sentence. Experiment 1 manipulates number mismatch, while Experiment 2 manipulates gender mismatch in French object relative clauses. Results show that number and gender mismatch increase production errors and reaction times, a finding that extends attraction effects in a new agreement dependency never investigated before, object-past participle agreement in French. Taken together, the two forced-choice experiments and the two self-paced reading experiments on number and gender object-past participle agreement in French, show that similarity-based
interference underlies mismatch effects both in production and in comprehension. However, I argue that while similarity plays a role in the process of controller selection in agreement production, it plays a role in the process of structure building in the comprehension of complex structures such as object relatives.

The experimental and theoretical work reported in this chapter is the result of years of collaborations with Julie Franck. Experiments 1-4 are reported in Villata and Franck (submitted).

1.4.4 Chapter 5: The mechanisms underlying interference effects

Chapter 5 starts from the observation of Chapter 4 that similarity-based interference effects can also be the result of encoding interference, possibly as a consequence of a mechanism of feature overwriting. A direct implication of this observation is that similar interference effects should also be observed when no agreement retrieval cue is present at the verb, since feature overwriting is supposed to arise independently from the presence of a retrieval cue, unlike cue-overload. The main goal of this Chapter is to test this implication with the aim to determine whether similarity-based interference effects are at play at retrieval, encoding or both. Experiment 1 tests for gender agreement in Italian object relative clauses with a self-paced reading procedure. Since in Italian the verb does not agree in gender with its subject, any observed mismatch effect (i.e., faster reading times and/or higher accuracy comprehension) would provide strong evidence in support of encoding interference, given that gender is not a retrieval cue in this language. Experiment 2 tests for number agreement in English object relative clauses through a self-paced reading procedure. In this experiment, the presence of an overt agreement cue on the verb is manipulated, by contrasting present tense verbs, which overtly agree in number with the third singular person, and past tense verbs, which do not carry distinctive number morphology. Results attest to a gender mismatch facilitation in Italian, both in comprehension accuracy and in reading times at the relative verb region. A number mismatch facilitatory effect was found in English in comprehension accuracy, regardless of the presence of an overt retrieval cue on the verb. However, a tendency for faster reading times in the number mismatch condition was observed in English only in the condition with a retrieval cue on the verb. Finally, online effects for both experiments were driven by particular slow trials. Results
from both experiments are thus consistent with interference arising at encoding, while only weak evidence for effects arising at retrieval is observed (online results in English only). These results stand in contrast with studies emphasizing the key role of retrieval interference in the resolution of long distance dependencies, as well as with studies reporting no effect of feature mismatch in grammatical agreement dependencies (e.g., Wagers et al. 2009, Dillon et al. 2013, Lago et al. 2015). I conclude by discussing how encoding effects can be implemented respectively in the ACT-R model of sentence processing (Lewis and Vasishth 2015) and in a self-organized sentence processing model (SOSP; e.g., Tabor and Hutchins, 2004).

The experimental and theoretical work reported in this chapter is the result of the collaboration with Julie Franck and Whit Tabor. Experiments 1 and 2 are reported in Villata, Tabor and Franck (submitted).
Chapter 2

Intervention effects and Relativized Minimality: New experimental evidence from graded judgments

According to Featural Relativized Minimality, the local relation between an extracted element and its trace is disrupted when it crosses an intervening c-commanding element whose morphosyntactic featural specification matches the specification of the elements it separates. This approach naturally leads to a system able to capture various degrees of deviance: the relative acceptability of an intervention configuration will vary as a function of the total, partial or zero feature overlap between the intervener and the target. Configurations involving a lesser degree of feature overlap should thus be more acceptable than sentences involving a higher degree of overlap. In this Chapter, I investigate the relative acceptability of wh-islands involving different degrees of criterial and non-criterial feature overlap. With this aim, I present three acceptability judgment experiments in which four configurations of criterial features overlap are systematically tested on a large set of linguistically naïve participants and with different methods of data collection. Results from the three experiments are highly consistent: predictions from Featural Relativized Minimality are globally borne out, except for the configuration involving two lexically restricted wh-elements (Complex Identity), for which an explanation in terms of grammar and processing is sketched out.

2.1 Introduction

In Chapter 1 (section 1.2), I discussed how Featural Relavized Minimality ranks structures from a minimum to a maximum of distinctiveness between X, the extracted element, and Z, the intervening element. I also discussed that features are arranged in a hierarchy such that strong degradation occurred when the overlap is on features triggering movement on their own (criterial Type I features), as it is the case for wh-islands, while an overlap on non-criterial Type
II features (i.e., features that trigger movement only when accompanied with a Type I feature) does not undermine the grammaticality of the sentence, as it is the case for object relatives. Let us consider again the configurations discussed in Chapter 1, reported here again for convenience.

(1) | X   | Z     | Y   |
---|------|-------|-----|
| a. Identity: | +A        | +A | +A |
| b. Inclusion:  | +A,+B     | +A | +A,+B |
| d. Disjunction: | +A       | +B | +A |

These configurations represent three possible patterns of relation between the featural specification of the intervener (Z) and the featural specification of the extractee (X): whenever the featural specification of the intervener is identical to the featural specification of the extractee (1a) an intervention effect arises determining a violation of fRM such that no local relation can be established between the target X and its trace Y, and the sentence is ruled out. The opposite case is Disjunction (1c): if the potential intervener Z has a disjoint specification with respect to the target, no intervention effect arises, and the local relation between X and Y can successfully be established. The Inclusion case (1b) represents an intermediate case: when the featural specification of Z is properly included in the specification of X and Y, the feature match is partial. Given that the principle states that a violation of locality arises in case of full match of feature (cfr. criterion (10b), section 1.2.2), Inclusion satisfies the principle of fRM. However, under the assumption that the amount of overlap provides a measure of the degree of deviance, we may expect that this intermediate situation gives rise to a violation which is felt as intermediate between the full acceptability of the disjunction case and the stronger ill-formedness of the identity case. In other words, we may interpret the second criterion of fRM “Z fully matches the specification in morphosyntactic features of X” along the following lines:

(2) a. Full match: strong violation
| b. Partial match: weak violation
| c. No match: well-formedness

In this view, the grammar does not simply determine a bifurcation between “grammatical” and “ungrammatical”, but it assigns a more fine-grained degree of grammaticality, with three values, as in (2). The system remains nonetheless discrete, as only a small number of set-
theoretic relations between relevant features are involved, with the traditional binary divide replaced by a more fine-grained discrete gradation.

### 2.2 Plan of the study

In order to systematically explore the effect of an overlap on criterial Type I feature, as well as the modulation in acceptability rates ensured by non-criterial Type II features, I manipulated the features overlap between the extractee and the intervener giving rise to four conditions illustrated in (3i). From now on, I will refer to conditions in (3i) as Intervention conditions, since they always involve the extraction of a wh-element over an intervening wh-element. These configurations instantiate four different set-theoretic configurations that I already discussed in Chapter 1 and that I briefly summarize here: (3ai) instantiates Bare Identity, since the identity of the featural specification between X and Z involves one single feature; (3bi) instantiates Inverse Inclusion, with the specification of Z properly including the specification of X; (3ci) instantiates Inclusion, with the specification of X properly including the specification of Z; finally, (3di) instantiates Complex Identity, in which the identity of the featural specification between X and Z involves more than one feature.

With the aim of controlling for any effect of lexical restriction independent of intervention, the Intervention conditions are compared to four baseline conditions without intervention (3ii) (see Atkinson et al. 2015 for a similar design in English). Note that sentences with intervention involve two other sources of variability that the No intervention conditions did not control for (see Sprouse et al., 2016). One is the presence of a long-distance dependency: Intervention conditions contain a long-distance dependency that No Intervention conditions do not. Hence, the current design does not allow us to determine whether the configuration that creates the weakest intervention effect (Complex identity) still shows any intervention effect or whether it only shows a long-distance dependency effect. Thus, I will capitalize on the relative rates across the four configurations of intervention, under the reasonable assumption that the effect of long-distance dependency is equal across them. The second source of variability in target conditions lies in the animacy of the two wh-elements: the inanimate objects used in the Intervention conditions were paired with animate main clause subjects in the No Intervention conditions.
Animate subjects were introduced in the latter in order to avoid introducing an independent source of difficulty (e.g., Traxler, Pickering, and McElree, 2002; Lowder, Gordon 2012).

(3)a. **Bare Identity:**

i. *Intervention:*

Qu’est-ce que tu te demandes [qui i a résolu j] ?

what do you wonder who solved?

[+Q] [+Q]

ii. *No Intervention:*

Qui j se demande [qui i a résolu ce problème]?

who wonders who solved this problem?

[+Q] [+Q]

b. **Inverse Inclusion**

i. *Intervention:*

Qu’est-ce que tu te demandes [quel étudiant i a résolu j] ?

what do you wonder which student solved?

[+Q] [+Q, +N]

ii. *No Intervention:*

Qui j se demande [quel étudiant i a résolu ce problème]?

who wonders which student solved this problem?

[+Q] [+Q, +N]

c. **Inclusion**

i. *Intervention:*

Quel problème te demandes-tu [qui i a résolu j] ?

which problem do you wonder who solved?

[+Q, +N] [+Q]

ii. *No Intervention:*

Quel professeur j se demande [qui i a résolu ce problème]?

which professor wonders who solved this problem?

[+Q, +N] [+Q]

d. **Complex Identity**

i. *Intervention:*

Quel problème te demandes-tu [quel étudiant i a résolu j] ?

which problem do you wonder which student solved?

[+Q, +N] [+Q, +N]

ii. *No Intervention:*

Quel professeur j se demande [quel étudiant i a résolu ce problème]?

which professor wonders which student solved this problem?

[+Q, +N] [+Q, +N]
In this design, the predictions of fRM bear on the strength of the intervention effects defined by comparing target structures with intervention (i.e., the wh-islands) to their corresponding baselines without intervention, rather than on raw acceptability rates in the target structures as proposed by the classical approach. If these sentences are presented out of context, the theory predicts stronger intervention effects for Bare Identity, where the intervener fully matches the featural specification of the extracted argument, than for Inclusion where it only partially matches it. As for Complex Identity, an unqualified version of fRM predicts it to be on a par with Bare Identity, since it also involves an intervener whose features fully match those of the extractee (but see below for an important qualification). Finally, Inverse Inclusion is also predicted to show a strong intervention effect similar to that for Bare Identity since its featural set fully matches that of the extractee (even though it is actually larger). The predictions are summarized in (4).

(4) Predictions of fRM for sentences out of context, if lexical restriction is a formal property defining feature sets, where “>” means “stronger intervention effect than” and “=” means “same intervention effect as”:

\[ \text{Bare Identity} = \text{Inverse Inclusion} = \text{Complex Identity} > \text{Inclusion} \]

However, as already discussed in Chapter 1 (section 1.2), the prediction that Complex Identity should pattern alike with Bare Identity and Inverse Inclusion is already known to be invalidated by informally gathered judgments which suggest that Complex Identity is better than both Bare Identity and Inverse Inclusion. The problem was noticed on the basis of the Italian equivalents of such structures in Rizzi 2011, and was tackled in the following way. We have seen that there are good reasons to assume that bare and lexically restricted wh-elements have distinct landing sites (see Chapter 1, section 1.1), so that the map of the left periphery of the clause involves a pure [+Q] attracting head, and a (higher) attracting head with the complex specification [+Q, +N]. Suppose now that a [+Q, +N] wh-element like which student has both movement options: it can be attracted by the [+Q, +N] head, and also by the simple, unqualified [+Q] head, as bare wh-elements (the opposite would not be true: so, a bare wh-element cannot be attracted by a [+Q, +N] head because it does not match its specifications, so that it only has the option of moving to a [+Q] head). In other words, we can assume the following (adapted from Rizzi 2011: 230, where the option is expressed in a framework assuming the topical feature of D-
linked phrases to be the crucial ameliorating factor; the approach is immediately translatable into a system assuming the lexical restriction to be the crucial factor):

(5) which N can be attracted both by [+Q, +N] and by [+Q]

I continue to assume that only the attracting features are seen in the computation of fRM. So, Complex Identity (3di) permits a representation like (5), derived by moving which student to Spec of [+Q] in the embedded clause, and which book to Spec of [+Q, +N] in the main clause:

(6) Which book do you wonder [which student __ could buy __ ]?

\([+Q, +N] \quad [+Q]\)

Under this analysis, the case of Complex Identity would reduce to a case of Inclusion. In fact, the judgment on the Italian equivalents assumed in Rizzi 2011 puts the two cases about on a par.\(^1\) In sum, if proviso (5) is adopted, the set of predictions (4) changes as follows:

(4') Predictions of fRM for sentences out of context, if lexical restriction is a formal property defining feature sets, and proviso (5) is adopted, where “>” means “stronger intervention effect than” and “=” means “same intervention effect as”:

\[
\text{Bare Identity} = \text{Inverse Inclusion} > \text{Inclusion} = \text{Complex Identity}
\]

I would now like to find out what the formal design described above has to say on this set of predictions. The study also aims at addressing experimentally the precise factor(s) that distinguish(es) bare and complex wh-elements (i.e., what vs. which NP). The approach I adopt

\(^1\) One could wonder why, if proviso (5) holds, a symmetric proviso could not hold for bare wh-elements, as in (5'), which would incorrectly permit bare elements to be (marginally) extracted:

(5') What can be attracted both by [+Q, +N], and by [+Q]

The answer is that the mechanics of the system disallows (5') while allowing (5): what is uniquely specified as [+Q], hence it could not be attracted by the complex head [+Q, +N], requiring also the presence of a lexical restriction. On the other hand, a complex phrase is specified [+Q, +N], hence it meets the attraction conditions of both a complex head [+Q, +N], and of a simple head [+Q].
here identifies the crucial factor as a purely formal property, namely the presence of a lexical restriction in the wh-phrase. However, another influential approach advocates an interpretive property, namely the D-linked character of the wh-phrase, i.e., the fact that the range of the variable is a presupposed set, pre-established in the discourse and salient. The formal and interpretive properties often co-occur, in that lexically restricted wh-phrases tend to be interpreted as D-linked (and some wh-elements requiring a lexical restriction, such as which in English, are assumed to enforce the D-linked interpretation), whereas bare wh-phrases tend to be interpreted as non D-linked. But the two properties do not necessarily coincide: as Pesetsky 1987 pointed out in his seminal work introducing D-linking, a bare wh-element can be made D-linked contextually. For instance, the sentence *Who did John talk to?* uttered out of context does not presuppose a specific set of potential interlocutors, i.e., it is not D-linked. However, it does presuppose such a set in a context like the following: *Under such circumstances, John could have talked to Bill, Peter or Mary; in the end, who did he talk to?*

In order to disentangle the role of lexical restriction and that of D-linking, the present study explores whether the presence of a context modulates the intervention effect in the four configurations described in (3). Two sets of predictions stand out. The first one bears on the comparison between the four configurations. If the feature of D-linking is responsible for the modulation of intervention effects in wh-islands, the four structures, once in context, are expected to be on a par since they would all involve a configuration of Inclusion [+Top, +Q] … [+Q], Top being the feature associated with D-linking (under the natural extension of proviso (5) where D-linking is the ameliorative factor, following Rizzi 2011). If, in contrast, it is the formal feature of lexical restriction that underlies the modulation of intervention effects, then the predictions are those formulated in (4’). The second prediction bears on the effect of context for each of the individual configurations. If the feature of D-linking is responsible for the modulation of intervention effects in wh-islands, then both Bare Identity and Inverse Inclusion should show a stronger intervention effect out of context than in context (given that in context they reduce to a case of Inclusion), while the intervention effect of Inclusion and Complex Identity should remain unchanged (under proviso (5)), since they are cases of Inclusion both in context and out of context. If, in contrast, the feature of lexical restriction is responsible for the modulation of intervention effects, then the ratings for each of the four configurations should be identical in the presence and in the absence of context, since their featural specifications are identical in context and out of context.
In what follows, I present three acceptability judgment experiments. In Experiment 1, participants are asked to judge the acceptability of sentences presented out of context on a 7-points Likert scale. In Experiment 2, a large sample of binary acceptability judgments is collected on the same structures, in order to see if results obtained in Experiment 1 are replicable. Finally, in Experiment 3 each sentence is preceded by a context story in order to explore the role of D-linking in modulating acceptability judgments.

2.3 Experiment 1: Graded judgments on wh-islands

2.3.1 Method

Participants. Forty students at the University of Geneva participated in the experiment for course credit. All participants were between the ages of 18 and 26, and were native speakers of French. All participants were naïve to the purposes of the study.

Material. Two variables were manipulated in a 2x4 design: (1) Intervention (Intervention vs. No Intervention), and (2) Structure (Bare Identity, Inverse Inclusion, Inclusion, Complex Identity). The four levels of Structure were obtained by manipulating the lexical restriction of the extracted wh-object and the intervening wh-subject. The two variables were part of a fully crossed design with 8 experimental conditions and were manipulated within participants. Stimuli consisted of 4 groups of sentences distributed across the 8 experimental conditions. All participants saw all stimuli. An example set, translated into English, is reported in Table 2.1.²

² For convenience, I used the same labels for the No Intervention and the Intervention configurations. For instance, the term Inclusion for the No Intervention condition refers to the fact that the featural specification of the second wh-element is included in the featural specification of first wh-element, even though none of the wh-elements moves across the other one, as it is the case for the Intervention conditions.
Table 2.1. English translation of the French sentences in the 8 experimental conditions.

<table>
<thead>
<tr>
<th>Structure</th>
<th>No Intervention</th>
<th>Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bare Identity</td>
<td>Who wonders who solved this problem?</td>
<td>What do you wonder who solved?</td>
</tr>
<tr>
<td>Inverse Inclusion</td>
<td>Who wonders which student solved this problem?</td>
<td>What do you wonder which student solved?</td>
</tr>
<tr>
<td>Inclusion</td>
<td>Which professor wonders who solved this problem?</td>
<td>Which problem do you wonder who solved?</td>
</tr>
<tr>
<td>Complex Identity</td>
<td>Which professor wonders which student solved this problem?</td>
<td>Which problem do you wonder which student solved?</td>
</tr>
</tbody>
</table>

In the Intervention condition, the extracted wh-element was always of the form *qu’est-ce que* (*what*) when it was bare, and *quel NP* (*which NP*) when it was restricted. French allows several options for the wh-element corresponding to the English *what*, namely the in-situ *quoi* (an option that we discarded as we wanted a long-distance dependency), the moved *que* (which requires subject-verb inversion, e.g., *Que te demandes-tu qui a construit?*) and the moved *que* accompanied by the interrogative form *est-ce que* (pronounced /esk/) which is (synchronously) a kind of overt question marker and prevents the need of the subject-verb inversion. The *est-ce que* marker was used for sentences with bare objects (e.g., *Qu’est-ce que tu te demandes qui a lu?*), while the subject clitic inversion was used for sentences with lexically restricted objects (e.g., *Quel livre te demandes-tu qui a lu?*). This choice was motivated by informal judgments of naturalness gathered among native French speakers. In the No Intervention conditions, I avoided the use of *est-ce que* as it would have involved the use of the “que to qui” rule (‘que’ becomes ‘qui’ in this context to alleviate the ECP violation in case of subject movement), which for some speakers is not fully natural.

Half of the experimental sentences contained *demander* (*wonder*) as main verb, whereas the other half contained *savoir* (*know*). All sentences containing *demander* were affirmative while only half of the sentences containing *savoir* were affirmative. As a result, there were 24 affirmative sentences and 8 negative sentences. The extracted wh-element was always inanimate, whereas the intervening wh-element was always animate. Eighty-eight grammatical filler sentences were added to the experimental items in order to introduce some variability in the structures. The fillers were various forms of questions (e.g., extractions from a declarative,
simple questions, and additional embedding of the kind *Which problem do you think that Mary thinks that we could solve?*).

**Procedure.** The experiment was programmed with E-prime. Each sentence was presented on a computer screen one at a time. Participants were tested individually in experimental booths and asked to judge for the acceptability of the sentences on a 7-point Likert scale (1 corresponding to a totally unacceptable sentence and 7 to a perfectly acceptable sentence) by pressing one of the seven numbered buttons on the keyboard. Participants were first shown 3 examples of sentences (one corresponding to a totally unacceptable sentence, one to a nearly acceptable sentence and one to a perfectly acceptable sentence) and their respective ratings (1, 4 and 7), none of which involved islands. They were then presented with 10 training items in order to familiarize themselves with the Likert scale. There was no time constraint on responses. Three short pauses were administered during the task. The whole session lasted about 30 minutes.

### 2.3.2 Results

**Data analyses.** The whole data set consisting of 1280 data points was analysed without excluding any outliers. Data were analysed with mixed-effects models estimated with the lmerTest package (http://www.cran.r-project.org/web/packages/lmerTest//lmerTest.pdf) in the R software environment (R Development Core Team, 2011). The 2x4 model has Intervention and Structure as fixed factors and involves random intercepts for subjects and items. The Satterthwaite approximation for degrees of freedom was used to estimate p-values. In order to explore the predictions of fRM bearing on the comparison across target structures involving intervention (i.e., wh-islands) controlled for their corresponding baselines without intervention, I ran six additional lmer models with, as fixed factors, Intervention and Structure Comparison, a 2-levels factor representing the six relevant contrasts between structures:

Model 1: Intervention (Intervention vs. No intervention) * Structure comparison 1 (Bare Identity vs. Inverse Inclusion)

Model 2: Intervention (Intervention vs. No intervention) * Structure comparison 2 (Bare Identity vs. Inclusion)

Model 3: Intervention (Intervention vs. No intervention) * Structure comparison 3 (Bare
Identity vs. Complex Identity)

Model 4: Intervention (Intervention vs. No intervention) * Structure comparison 4 (Inverse Inclusion vs. Inclusion)
Model 5: Intervention (Intervention vs. No intervention) * Structure comparison 5 (Inverse Inclusion vs. Complex Identity)
Model 6: Intervention (Intervention vs. No intervention) * Structure comparison 6 (Inclusion vs. Complex Identity)

If the key differences between the four structures predicted by fRIM are tightly linked to intervention, that is, if they attest to intervention effects, these differences should not show up in the corresponding baseline conditions with no intervention, or they should at least be reduced. That is, the predictions of the theory bear on the interactions between Structure Comparison and Intervention: if the theory predicts a stronger intervention effect in structure A than in structure B, this prediction will translate into a bigger difference between A and B in the Intervention condition than in the No intervention, baseline condition. Under prediction (4), Inclusion is expected to show the weakest intervention effect; thus, Models 2, 4 and 6 containing Inclusion are all expected to show an interaction, with a difference between Inclusion and the comparison structure (Bare Identity in Model 2, Inverse Inclusion in Model 4, Complex Identity in Model 6) being greater in the Intervention condition than in the baseline.

If proviso (5) adopted in prediction (4') is correct, Complex Identity is expected to show a weaker intervention effect than Bare Identity and Inverse Inclusion, which will manifest by way of a significant interaction in Models 3 and 5. Finally, similar intervention effects in Complex Identity and Inclusion should be found under (4'), which should translate in terms of a lack of interaction in Model 6.

**Results.** Figure 2.1 reports the mean acceptability scores and standard errors for the four sets in the target conditions with intervention and the baseline controls. Acceptability rates are significantly higher in the baseline conditions without Intervention ($M = 5.9$) than in the conditions with Intervention ($M = 3.2$) ($F = 1379.84, p < .001$), and they significantly differ across the four structures ($F = 4.55, p = .003$). The significant interaction between Intervention and Structure ($F = 47.09, p < .001$) shows that the effect of Structure is not the same for the conditions with intervention and for their corresponding baselines.
**Figure 2.1.** Plot of the 8 individual structures on the 7-point Likert scale for Experiment 1.

A summary of the interaction tests for the six models with Intervention and Structure Comparison as factors is reported in Table 2.2. Except for Model 1, all other interactions reached statistical significance, showing that the effect of structure is greater in the conditions with intervention than in their baseline controls. In sum, results show the following pattern of intervention effects (where “>” means “stronger intervention effect than” and “=” means “same intervention effect as”):

**Bare Identity = Inverse Inclusion > Inclusion > Complex Identity.**

<table>
<thead>
<tr>
<th>Variable</th>
<th>$F$ value</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1: Bare Identity vs. Inverse Inclusion</td>
<td>0.39</td>
<td>.531</td>
</tr>
<tr>
<td>Model 2: Bare Identity vs. Inclusion</td>
<td>30.56</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Model 3: Bare Identity vs. Complex Identity</td>
<td>95.02</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Model 4: Inverse Inclusion vs. Inclusion</td>
<td>35.0</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Model 5: Inverse Inclusion vs. Complex Identity</td>
<td>99.98</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Model 6: Inclusion vs. Complex Identity</td>
<td>16.799</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
2.3.3 Discussion

Experiment 1 was designed to systematically assess predictions from fRM, recalled here for convenience:

(4') Predictions of fRM for sentences out of context, if lexical restriction is a formal property defining feature sets, and proviso (5) is adopted (”>” means “stronger intervention effect than” and “=” means “same intervention effect as”):

Bare Identity = Inverse Inclusion > Inclusion = Complex Identity

The results thus validate the prediction that Bare Identity and Inverse Inclusion show the same intervention effect, which in turn is stronger than Complex Identity and Inclusion. However, Complex Identity turned out to be less sensitive to intervention effects than Inclusion, which was not expected under fRM. I return to this unexpected result in the General Discussion.

Before extending the discussion of these findings further, it appeared important to determine whether these results are replicable and generalizable. With this aim in mind, I ran the same experiment by increasing the number of items to 320, which allows us to investigate if results from Experiment 1 are replicable, which is always desirable in experimental research, and generalizable to configurations with the same structure but different lexical materials. Moreover, in the next experiment the 7-point Likert scale was substituted by a binary scale, in order to investigate if these effects are captured by binary categorical judgments, which is interesting given that many models of grammaticality are categorical.

2.4 Experiment 2: Binary judgments on wh-islands

2.4.1 Method

Participants. Twenty students from the University of Geneva participated in the experiment for course credit. All were between the ages of 18 and 26, and were native speakers of French. Participants recruited for this experiment had not taken part in Experiment 1.
Materials. The same variables manipulated in Experiment 1 were manipulated in the same design: (1) Intervention (Intervention vs. No Intervention), and (2) Structure (Bare Identity, Inverse Inclusion, Inclusion, Complex Identity). Stimuli consisted of 40 groups of sentences distributed across the 8 experimental conditions. Experimental sentences were intermixed with 320 fillers created by replacing the transitive embedded verb of experimental sentences (e.g., *solve*) with an intransitive verb (e.g., *sleep*). These fillers served as experimental sentences for another experiment not reported here. The 640 resulting items (320 experimental sentences and 320 fillers) were split into two between-subjects lists containing 320 items each.

Procedure. The experiment was programmed with LimeSurvey (http://www.limesurvey.com/), a free and open-source software that makes it possible to create online surveys. The questionnaire was administered individually in experimental booths. Participants were asked to judge the acceptability of each sentence on a binary scale by clicking on ‘yes’ if the sentence was acceptable and on ‘no’ if it was unacceptable. Participants were first shown 9 example sentences (5 acceptable and 4 unacceptable) and their respective ratings, none of which involved islands. LimeSurvey displayed all sentences on the computer screen, one after the other. The whole session lasted about 30 minutes.

2.4.2 Results

Data Analysis. The whole data set consisting of 6400 data points was analysed without excluding outliers. Generalized mixed effects models for binomial distribution (GLME) were estimated in the R software environment (R Development Core Team, 2011). The fixed factors of the model are Intervention and Structure. The model involves random intercepts for subjects and items. Likelihood ratio tests were run in order to test for main effects and interactions. The additional theory-based analyses involving structure comparison were the same as described in Experiment 1.

Results. Figure 2.2 illustrates the mean acceptability scores and standard errors for the four configurations in the target conditions with intervention and the baseline controls without intervention. The acceptability rates are significantly higher in the No Intervention conditions ($M = 0.95$) than in the Intervention conditions ($M = 0.28$) ($\chi^2(1) = 2007.2$, $p < .001$), and they
significantly differ across the four structures ($\chi^2 (3) = 218.92, p < .001$). The significant interaction between Intervention and Structure ($\chi^2(3) = 178.87, p < .001$) reveals that the effect of Structure is not the same for the conditions with intervention and for the baselines.

**Figure 2.2** Plot of the 8 individual structures on the binary scale for Experiment 2.

A summary of the interaction tests for the six models with Intervention and Structure Comparison as fixed factors is reported in Table 2.3. Except in Model 1, all interactions were statistically significant, replicating results from Experiment 1. In sum, results show the following pattern of intervention effects (where “>” means “stronger intervention effect than” and “=” means “same intervention effect as”):

*Bare Identity = Inverse Inclusion > Inclusion > Complex Identity*

**Table 2.3.** Summary of the interaction tests for the six models in Experiment 2.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>SE</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1: Bare Identity vs. Inverse Inclusion</td>
<td>0.073</td>
<td>0.564</td>
<td>0.130</td>
<td>.896</td>
</tr>
<tr>
<td>Model 2: Bare Identity vs. Inclusion</td>
<td>3.137</td>
<td>0.487</td>
<td>6.436</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Model 3: Bare Identity vs. Complex Identity</td>
<td>4.509</td>
<td>0.493</td>
<td>9.137</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Model 4: Inverse Inclusion vs. Inclusion</td>
<td>3.248</td>
<td>0.460</td>
<td>7.056</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Model 5: Inverse Inclusion vs. Complex Identity</td>
<td>4.591</td>
<td>0.462</td>
<td>9.922</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Model 6: Inclusion vs. Complex Identity</td>
<td>1.716</td>
<td>0.339</td>
<td>5.052</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
2.4.3 Discussion

Results from Experiment 2 replicate those from Experiment 1 with a different task involving binary judgments and with an increased number of items, attesting to the robustness of the data. In particular, the finding that configurations involving a lexically restricted extractee (Inclusion and Complex Identity) are less sensitive to intervention effects than configurations involving a bare extractee (Bare Identity and Inverse Inclusion) was replicated, as predicted by fRM under proviso (5).

As discussed in the Introduction, two independent (but often co-occurring) factors could be responsible for the decrease of the intervention effect found for sentences involving lexically restricted extractees: one factor is lexical restriction, a structural feature, the other is the D(linked) character of the extractee. In order to assess the potential role of D-linking, I ran a third experiment in which each sentence was preceded by a context story, in order to make bare wh-phrases linked to the discourse (i.e., D-linked).

2.5 Experiment 3: D-linking or lexical restriction?

2.5.1 Method

Participants. Forty-nine students from the University of Geneva participated in the experiment for course credit. All were between the ages of 18 and 26, and were native speakers of French. Participants recruited for this Experiment had not taken part in Experiment 1 or in Experiment 2.

Materials. For this experiment the very same material as Experiment 1 was used, except that each question was preceded by a short context story. The context story was designed to promote D-linking by introducing a set of entities to which the wh-elements refer to. For each group of sentences two stories were created: one for the four No Intervention configurations and another one for the corresponding four Intervention configurations. Examples of context stories for one set of sentences are given in Table 2.4. Four lists were created such that each participant read a
story only once. Each list contained 30 sentences in total (8 experimental sentences and 22 fillers).

Table 2.4. Example of context story and the corresponding questions for Experiment 3.

<table>
<thead>
<tr>
<th>Story</th>
<th>Questions</th>
</tr>
</thead>
</table>
| **No Intervention** | | 1. Who is wondering who has solved this problem?  
2. Who is wondering which student has solved this problem?  
3. Which physicist is wondering who has solved this problem?  
4. Which physicist is wondering which student has solved this problem? |
| In the Physics department, there is a blackboard on which each month a physicist writes a problem to be solved by a student. This month a student has solved the problem only 2 hours after the problem was written on the blackboard. The professors are astonished! Among the physicists, Arnold knows that Frédéric has solved the problem, because he told him. But Pascal, another physicist, is not aware of this. According to him, the two students likely to have solved the problem are Frédéric and Nicolas. Pascal would really like to know who the author of the solution is and he will inquire! |  |
| **Intervention** | | 5. What did you wonder who solved?  
6. What did you wonder which student solved?  
7. Which problem did you wonder who solved?  
8. Which problem did you wonder which student solved? |
| You are a student in Mathematics. Last week your professor assigned to your class 5 problems to be solved. You have solved the first four problems, but the fifth one was so difficult that you did not hand it out. When you arrived in the class this morning, the professor has announced that only one student has been able to solve the fifth problem. You would really like to know who the genius that has been able to do it is! |  |

**Procedure.** The procedure was the same as the one used in Experiment 1, except that each context story was presented on the computer screen before the presentation of the question. Participants pressed the space bar to switch from the story to the question.

**2.5.2 Results**

**Data analyses.** One subject was excluded from the analysis since 75% of his scores were 2 standard deviations away from the mean of the population. Analyses were thus run on the 384
remaining data points. The same data analyses as Experiment 1 were conducted, with a first model involving Intervention and Structure followed by six models involving Intervention and Structure comparison. If the key factor underlying intervention effects is D-linking, all configurations will show the same intervention effect since, once in context, they all reduce to a configuration of Inclusion (i.e., [+Top,+Q]...[+Q]). This will translate into a lack of interaction between Intervention and Structure comparison in all six models.

**Results.** Figure 2.3 reports the mean acceptability scores and standard errors for the four configurations in the target conditions with intervention and the baseline controls without intervention. The acceptability rates are significantly higher in the conditions without Intervention ($M = 5.9$) than in the conditions with Intervention ($M = 2.9$) ($F = 549.80, p < .001$), and they significantly differ across the four structures ($F = 4.49, p = .004$). The significant interaction between Intervention and Structure ($F = 12.88, p < .001$) shows that the effect of Structure is not the same for the conditions with intervention and for the baselines.

![Figure 2.3. Plot of the 8 individual structures on the 7-points Likert scale for Experiment 3.](image)

A summary of the interaction tests for the six models with Intervention and Structure Comparison as fixed factors is reported in Table 2.5. The interactions in Models 3-6 were significant, showing that the intervention effect is stronger in Inverse Inclusion than in Inclusion and Complex Identity, that it is stronger in Inclusion than Complex Identity, and that it is stronger in Bare Identity than in Complex Identity. The lack of interaction in Model 1 shows
that the intervention effect in Bare Identity does not differ from that in Inverse Inclusion. The
interaction in Model 2 did not reach significance level, suggesting that the intervention effect
in Bare Identity is not statistically different from Inclusion, although a trend was found towards
a stronger effect in Bare Identity. In sum, results from individual models show the following
pattern (where “>” means “stronger intervention effect than” and “=” means “similar
intervention effect as”):

*Inverse Inclusion > Inclusion > Complex Identity*

*Inverse Inclusion = Bare Identity > Complex identity*

*Bare Identity = Inclusion*

**Table 2.5.** Summary of the interaction tests for the six models in Experiment 3.

<table>
<thead>
<tr>
<th>Variable</th>
<th>$F$ value</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1: Bare Identity vs. Inverse Inclusion</td>
<td>1.63</td>
<td>.203</td>
</tr>
<tr>
<td>Model 2: Bare Identity vs. Inclusion</td>
<td>1.794</td>
<td>.188</td>
</tr>
<tr>
<td>Model 3: Bare Identity vs. Complex Identity</td>
<td>19.252</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Model 4: Inverse Inclusion vs. Inclusion</td>
<td>6.75</td>
<td>.01</td>
</tr>
<tr>
<td>Model 5: Inverse Inclusion vs. Complex Identity</td>
<td>8.349</td>
<td>.004</td>
</tr>
<tr>
<td>Model 6: Inclusion vs. Complex Identity</td>
<td>8.776</td>
<td>.003</td>
</tr>
</tbody>
</table>

**2.5.3 Effect of context: Comparison between Experiments 1 and 3**

I now focus on the second prediction discussed in section 2.2, which bears on the effect of
context for each of the individual structures. As we have seen, if the feature of D-linking is
responsible for the modulation of intervention effects in wh-islands, then we expect both Bare
Identity and Inverse Inclusion to show a stronger intervention effect out of context than in
context, while no effect should be found for Inclusion and Complex Identity. If, in contrast, the
feature of lexical restriction is the crucial factor in the modulation of intervention effects, then
the ratings for each of the four configurations should be identical in the presence and in the
absence of context. With the aim of testing these predictions, I ran various models on the results of Experiments 1 and 3 merged together with Context as a between-participants factor. Figure 2.4 summarizes results from Experiment 1 and 3. A first general model with Structure, Intervention and Context as fixed factors was run, in order to test for the interactions involving Context. Results showed a marginally significant interaction between Context and Intervention ($F = 3.09, p = .07$), attesting to a tendency for stronger intervention effects for structures in context (No intervention: $M = 5.92$; Intervention: $M = 2.99$) than for structures without context (No intervention: $M = 5.91$; Intervention: $M = 3.24$). The interaction between Context and Structure was not significant ($F = 0.98, p = .40$), revealing that the effect of structure is not different when there is context from when there is no context. The three-way interaction between Structure, Intervention and Context was not significant ($F = 0.71, p = .55$).

**Figure 2.4.** Plot of the 8 individual structures on the 7-points Likert scale for Experiments 1 and 3. The first four structures on the left are conditions with No Intervention, while the four structures on the right are conditions with Intervention.

In order to test predictions about the effect of context for each structure, four separate models were run with Intervention and Context as factors. If Context modulates the intervention effect

---

3 Note that the design obtained by merging experiments 1 and 3 is unbalanced, in that there are more data points in the No Context condition than in the Context condition. However, mixed-effects models are robust for unbalanced designs (e.g., Baayen et al. 2008).
in a particular structure, an interaction is expected between Context and Intervention. Model 1 tested the interaction between Intervention and Context for Bare Identity, Model 2 for Inverse Inclusion, Model 3 for Inclusion and Model 4 for Complex Identity. Results from the four models are reported in Table 2.6. None of the models reached statistical significance, suggesting that the presence of context does not significantly modulate intervention effects. Model 3 is nevertheless marginally significant \((p = .074)\), showing that the intervention effect for the Inclusion configuration is actually stronger in the presence of a context (No intervention: \(M = 6.125\); Intervention: \(M = 3.25\)) than in the absence of context (No intervention: \(M = 5.93\); Intervention: \(M = 3.61\)).

### Table 2.6. Summary of the interaction tests for the four models in Experiments 1 and 3.

<table>
<thead>
<tr>
<th>Variable</th>
<th>F value</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1: Bare Identity out of context*Bare Identity in context</td>
<td>0.03</td>
<td>.853</td>
</tr>
<tr>
<td>Model 2: Inverse Inclusion out of context*Inverse Inclusion in context</td>
<td>0.83</td>
<td>.362</td>
</tr>
<tr>
<td>Model 3: Inclusion out of context*Inclusion in context</td>
<td>3.208</td>
<td>.074</td>
</tr>
<tr>
<td>Model 4: Complex Identity out of context*Complex Identity in context</td>
<td>0.732</td>
<td>.393</td>
</tr>
</tbody>
</table>

#### 2.5.4 Discussion

Results from Experiment 3 show some differences in the intervention effects of the four individual structures when sentences were presented in context. This is unexpected under the hypothesis that D-linking is the relevant factor in modulating intervention effects, since this hypothesis predicted that no difference should be found amongst the four structures (see section 1.2). Nevertheless, even though some of the structures showed different intervention effects with and without context, others did not, in contrast to what was found in Experiments 1 and 2, and in line with the D-linking hypothesis. I found that the intervention effect for Bare Identity did not differ from that for Inverse Inclusion nor from that for Inclusion. However, notice that the lack of difference between Bare Identity and Inverse Inclusion was also predicted by the lexical restriction hypothesis (4'). The finding that Bare Identity showed a similar intervention effect as Inclusion, however, was predicted by the D-linking hypothesis, but not by the lexical
restriction hypothesis, which predicted a weaker intervention effect for Inclusion (4’). Yet, closer inspection of the effect of context on the configuration of Inclusion shows that the lack of difference between Bare identity and Inclusion in the presence of context is actually not due to a reduced intervention effect for Bare Identity in context, as predicted by the D-linking hypothesis, but to an increased intervention effect for Inclusion in context. Given that none of hypotheses predicted that context should increase the intervention effect in Inclusion, I suggest that this data point is noise.

Finally, it is important to note that the finding that context did not modulate intervention effects (or at least not in any relevant way) bears on a null result, which may also be due to the fact that the context stories I used were not able to induce D-linking, a possibility that cannot be discarded (see Sprouse’s dissertation 2007 for a similar null result in Superiority violations).

2.6 General Discussion

A series of three experiments explored the predictions from featural Relativized Minimality on the acceptability of wh-islands, with an experimental design allowing us to control for variables independent of the critical variables assumed by the theory. The theory assumes that when two elements that should enter into a local relation are separated by an element matching the featural specification of the elements it separates, an intervention effect arises and the sentence is perceived as degraded. According to the version of the principle I have adopted here, the features that are taken into account by the system are morphosyntactic features attracting movement. When the intervener fully matches the morphosyntactic featural specification of the target (configurations of Identity and Inverse Inclusion), the structure is expected to be severely ill-formed; when the intervener matches the specification of the target only in part (configuration of Inclusion) a milder deviance should arise.

Capitalizing on theoretical work suggesting that lexical restriction is a feature triggering movement and on experimental evidence showing that it plays a role in modulating intervention effects in object relative clause comprehension in acquisition (Friedmann et al. 2009, Munaro 1999), I tested four set theoretic configurations: Bare Identity (both wh-elements are bare), Complex Identity (both wh-elements are lexically restricted), Inclusion (the extracted wh-element is lexically restricted but the intervening one is not) and Inverse inclusion (the extracted
wh-element is not lexically restricted but the intervener is). Experiment 1 involved a 7-point scale acceptability judgment procedure, Experiment 2 involved a binary acceptability procedure, and Experiment 3 was identical to Experiment 1 except that sentences were preceded by a discourse context. Results from the three experiments are highly consistent, replicating major effects with different methods of data collection.

2.6.1 Featural Relativized Minimality and weak islands

In line with the predictions of fRM, results from Experiments 1 and 2 show that when the feature match is complete, as in Bare Identity and Inverse Inclusion, the intervention effect is stronger than when it is partial, as in Inclusion and Complex Identity. As discussed in the Introduction, the weaker intervention effects observed for complex wh-extractees could either be due to their lexical restriction or to their D-linked character. Adding a short context story before the sentences allowed us to tease apart the role of these two factors in Experiment 3. A formally bare wh-element can indeed be made contextually D-linked if uttered in a context presupposing a set of entities which constitute the range of the wh-variable. Results from Experiment 3 show that the presence of a context did not contribute to reduce the intervention effects in any of the four configurations (it even tended to increase it in the Inclusion configuration, a result which remains unexplained), and the context did not fundamentally modify the gradient of intervention effects observed in Experiments 1 and 2. On that basis, I conclude that D-linking is not the relevant factor in modulating intervention effects, which rather appear carried by the formal syntactic property of lexical restriction of complex wh-elements.

Another relevant aspect of the data concerns the status of Complex Identity. All experiments showed a weaker intervention effect in the configuration of Complex Identity. Proviso (5) involved the assumption that Complex Identity can always admit a structural analysis with the lower complex wh-phrase attracted by a bare [+Q] head, as in (6). Therefore, if the featural specification which is taken into account is the one actually attracting movement, the case of Complex Identity can reduce to Inclusion, and thus display a weaker intervention effect than Bare Identity and Inverse Inclusion. The experimental data confirm these predictions. Nevertheless, the three experiments consistently show an even weaker intervention effect for
Complex Identity than for Inclusion, in line with what was also observed in English (Atkinson et al. 2015). One way to address these data would be to amend the system of fRM in order to capture this case as well. One possibility would be to broaden the class of features taken into account in the calculation of the constraint, currently taken to be morphosyntactic features triggering movement, to also include the lexical features expressed by the lexical restriction. Consider whatever featural property that defines the lexical noun problem and differentiates it from the lexical noun student in Which problem do you wonder which student solved? If such features are taken into account in the calculation of locality, the set-theoretic relation between which problem and which student would become one of Intersection: the two phrases would have in common [+Q] and [+N], and would differ in all the lexical features which differentiate problem from student. This possible revision gains some plausibility in view of the fact that the set-theoretic relation of Intersection was shown to enhance comprehension of object relatives in children in Belletti et al. 2012 in comparison to Inclusion. Nevertheless, the extension of the class of relevant features beyond the class of morphosyntactic features triggering movement raises problems elsewhere, for reasons discussed by Belletti et al. who argued in detail for a restrictive view of the feature calculus for locality.

The possibility that a wider range of features enters into the metric of similarity between the moved element and the intervener is actually closely aligned with recent theories of memory interference in sentence comprehension. In the next section, the relevant characteristics of these theories are sketched out in order to evaluate the extent to which they may account for our findings.

### 2.6.2 Cue-based memory models and similarity-based interference

A core feature of weak islands is that they contain two wh-elements that have to be integrated with the verb. Hence, when the relevant verb is reached, the representations of the arguments must be retrieved from memory in order to assign them their thematic roles and derive the proper interpretation. Content-addressable memory models provide a detailed analysis of how difficulties in parsing long-distance dependencies arise from constraints from memory retrieval mechanisms (e.g., Gordon et al. 2001, 2004, Lewis et al. 2006, McElree 2000, Van Dyke and McElree 2006). According to this framework, memory retrieval is driven by cues, which inform
the parser about the features identifying the element that it is looking for and distinguishing it from other irrelevant representations in memory. Cues enable direct access to the relevant representation, i.e., the to-be-retrieved element, without the need to search through all the stored representations, which ensures a rapid recovery of relevant information stored in memory. However, rapidity comes at a cost, and there are well-understood deficiencies of this type of retrieval mechanism. A cue-driven operation can fail to recover the intended representation if the retrieval cues do not sufficiently overlap with the features with which the representation has been encoded into memory. Even when retrieval cues do sufficiently overlap with the representation, retrieval may still fail if those cues also match, even partially, the contents of other items in memory. This condition is known as similarity-based interference (e.g., Gordon et al. 2001, Nairne 2002, Lewis et al. 2006, McElree 2006). Early reports that the similarity between an element to be retrieved from memory and an intervening element caused increased reading times and increased comprehension errors came from studies on object relative clauses. These studies showed that reducing the similarity between the subject and the object, like when the subject is pronominalized while the object is a full NP, reduced processing times at the verb (e.g., Gordon et al. 2001, 2004). Similarity-based interference has also been attested in the processing of subject-verb long-distance dependencies: lower acceptability and longer reading times at the verb have been observed for cases in which the syntactic relation between the subject and the verb was interrupted by a relative clause containing an intervening grammatical subject, while no such effects have been observed when the interpolated material contained an intervening NP occupying an object position (e.g., Van Dyke and Lewis 2003). Interestingly, other studies showed that semantically similar distractors could also generate interference (e.g., Van Dyke 2007).

The finding that both syntactic and semantic cues play a role in determining interference provides us with a possible explanation for the weaker intervention effect observed for Complex Identity as compared to Inclusion. This could be due to the greater semantic distinctiveness, and therefore the lower similarity of lexically restricted wh-elements. Indeed, although in Complex Identity both wh-elements are identical regarding their syntactic features (i.e., [+Q, +N]), they are endowed with rich bundles of semantic features which all contribute to increasing their distinctiveness. If we consider a sentence like Which problem do you wonder which student solved? it is highly plausible that the retrieval of the extracted wh-element (which problem) is eased by the fact that problems can be solved, therefore providing a good semantic
match to the object retrieval process triggered by the verb, whereas students cannot be solved, therefore providing a poor match, and reducing interference.

This hypothesis is supported by a series of studies conducted by Hofmeister et al. (2007, 2013), in which they found that increasing the syntactic and semantic complexity of the extracted element eases its retrieval and increases the acceptability of the sentence. For instance, using cleft constructions (e.g., It was a(n) (alleged Venezuelan) communist who the members of the club banned from ever entering the premises), they observed faster reading times after the embedded verb banned in sentences requiring the retrieval of complex objects (an alleged Venezuelan communist) as compared to those requiring the retrieval of simple objects (a communist) (Hofmeister et al. 2007). Interestingly, faster reading times after the verb coincided with slower reading times at the NP region (length being controlled). The authors took this result as indicative of a deeper encoding of complex NPs (an alleged Venezuelan communist vs. a communist), and argued that deeper encoding allowed for easier retrieval by providing distinctive retrieval cues.

In sum, psycholinguistic evidence supports the hypothesis that semantic richness plays a role in the processing of sentences involving memory retrieval, by reducing cue overlap and therefore similarity-based interference. The weaker intervention effect found for wh-islands containing lexically restricted wh-extractees may derive from the same principle of cue-based, direct access memory retrieval. In this view, interference is a function of the degree of similarity between the extractee and the intervener, where both syntactic and semantic features come into play in the similarity metric. I discuss these possible developments within and outside the system of fRM in Chapter 3 (on the comparison and the prospects for an integration between fRM and memory-based models see also Belletti and Rizzi 2013; Santos, 2011).

2.6.3 Relativized Minimality and Superiority

It has been suggested that Superiority could be reduced to a case of intervention (ultimately RM or analogous principles in minimalist approaches), in which case the ungrammaticality of a sentence of the kind of (7a) lies in the movement of what across a similar element who (Chomsky 1995). If so, we would expect lexical restriction to affect superiority violation configurations in the same way it affects configurations of intervention. However, this does not
seem to be the case, as some recent findings by Hofmeister et al. 2013 attested. In four experiments (two acceptability judgment and two self-paced reading experiments), the authors explored the role of lexical restriction in structures involving Superiority violations (Pesetsky 2000). They contrasted 4 conditions, illustrated in (7).

(7)  a. Mary wondered what who read.
     b. Mary wondered which book who read.
     c. Mary wondered what which boy read.
     d. Mary wondered which book which boy read.

In line with our results, the condition of Complex Identity (7d) always gave rise to the highest acceptability scores, whereas the condition of Bare Identity (7a) generated the lowest scores. However, the two intermediate conditions showed a different profile to ours. In particular, if we focus only on Hofmeister et al.’s acceptability judgment experiments (leaving aside the self-paced reading experiments which are less directly comparable to our acceptability judgments experiments), their first experiment showed a reverse profile to ours, with what appears like Inverse Inclusion (7c) showing higher scores than what appears like Inclusion (7b), while their fourth experiment showed the two conditions to be on a par. Interestingly, an increase in acceptability ratings for condition (7c) as compared to condition (7b) was also attested in Hofmeister et al. 2007 as well as in German (Featherston 2005). Hence, experimental evidence in English and German suggests the following pattern for Superiority violations (where “<” means “lower acceptability rates than”): Bare Identity < Inclusion < Inverse Inclusion < Complex Identity, with Inverse Inclusion being more acceptable than Inclusion, in contrast to the higher acceptability of Inclusion found in weak islands.

So, the issue arises of understanding the reversed effect of lexical restriction found for cases of multiple wh-questions (Inclusion < Inverse Inclusion in Hofmeister et al. 2013) and for wh-islands (Inclusion > Inverse inclusion in our acceptability results). In that regard, it is interesting

4 Notice that these results expressed in terms of acceptability rates easily translate in terms of intervention effect: high acceptability rates correspond to a weak intervention effect, while low acceptability rates correspond to a strong intervention effect. Hence, the sign ‘<’ is used in this section to express low acceptability rates, while the corresponding sign ‘>’ has been used in Experiments 1, 2 and 3 as expressing strong intervention effect.
to explore the consequences of the classical analysis of multiple questions in Chomsky 1981, according to which the wh-element which is in situ at surface structure moves to the left periphery covertly in the syntax of Logical Form, in order to be placed in the appropriate scope position as binder of a wh-variable on the appropriate level of representation at the interface with semantic interpretation. The respective Logical Forms of (7) are thus the following, after covert movement:

(8) a. Mary wondered who \_ what \_ read \_ i .
    b. Mary wondered who \_ which book \_ read \_ i .
    c. Mary wondered which boy \_ what \_ read \_ i .
    d. Mary wondered which boy \_ which book \_ read \_ i .

In this analysis, the locality violation determining the low acceptability of (7a) is determined by the covert movement of who crossing what moved to the left periphery in the overt syntax. If some version of this approach is on the right track, the reversal observed in Hofmeister et al.’s results (as well as in Featherston’s study on German) with respect to ours is immediately understandable because covert movement inverts the polarity between Inclusion and Inverse Inclusion: (7c) superficially looks like Inverse Inclusion, but if locality is computed on the covert movement of which boy, the relevant representation is (8c), a case of Inclusion; similarly, (7b) looks superficially like a case of Inclusion, but if the relation is computed on (8b) after covert movement of who, the case really instantiates Inverse Inclusion. The higher acceptability of (7c) over (7b) is thus expected, and in line with our results on extraction from weak island.\(^5\)

\(^5\) This solution, however, comes with a not negligible cost, as it would also predict a grammatical sentence like Who read what to be degraded: according to the analysis sketched above, the object would covertly move over the subject at the Logical Form, thus violating locality. Yet, the sentence is fully grammatical. In this dissertation, I will not tackle this problem, and I refer the reader to Mirdamadi, Villata, Shlonsky and Franck (submitted) for a thorough discussion and a tentative solution.
2.7 Conclusions

The use of a formal methodology for the gathering of grammaticality judgments shows a considerable heuristic value and potential for the enrichment of the empirical basis in a complex domain like the violation of weak islands. On the one hand, this methodology has provided detailed evidence confirming basic predictions of fRM, thus offering further support for this approach to intervention locality; on the other hand, it has opened new questions, namely in connection with the relatively high acceptability of Complex Identity and the analysis of Superiority violation constructions, which suggests that new ideas and lines of inquiry are worth exploring to further increase the empirical adequacy of the theory of intervention effects. The status of Complex Identity represents the starting point of Chapter 3.
2.8 Appendix

Here I report results of a 2x2x2 linear mixed effects models using the fixed factors Intervention, Lexical restriction of the extracted element (Wh1) and Lexical restriction of the intervening element (Wh2). For the sake of brevity (and since results obtained with this alternative method perfectly replicate those obtained in the 2x4 design), I report the alternative analysis for Experiment 1 only. Data were analysed with mixed-effects models estimated with the lmerTest package (http://www.cran.r-project.org/web/packages/lmerTest/lmerTest.pdf) in the R software environment (R Development Core Team, 2011). The fixed factors of the model are Intervention, Wh1 and Wh2. The model involves random intercepts for subjects and items. A summary of the main ratings in the 8 experimental conditions is reported in Table 2.A.1.

Table 2.A.1. Mean ratings on a 7 points Likert scale in the 8 experimental conditions.

<table>
<thead>
<tr>
<th></th>
<th>Wh1 = Non lexically restricted</th>
<th>Wh2 = Lexically restricted</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Intervention</td>
<td>6.1</td>
<td>6.2</td>
<td>6.2</td>
</tr>
<tr>
<td>Intervention</td>
<td>2.7</td>
<td>2.7</td>
<td>2.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Wh1 = Lexically restricted</th>
<th>Wh2 = Lexically restricted</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Intervention</td>
<td>5.9</td>
<td>5.3</td>
<td>5.6</td>
</tr>
<tr>
<td>Intervention</td>
<td>3.6</td>
<td>3.9</td>
<td>3.8</td>
</tr>
</tbody>
</table>

Results from mixed models analyses revealed a main effect of Intervention, attesting to significantly higher rates for sentences in the No Intervention condition than for sentences in the Intervention condition ($\beta = -2.67, t = -37.146, p < .001$). Results also attested to a significant triple interaction between Intervention, Wh1 and Wh2 ($\beta = 0.993, t = 3.456, p < .001$). Thus, subsequent models were conducted separately for the No Intervention and the Intervention conditions.

**Intervention conditions.** A summary of the fixed effects is reported in Table 2.A.2.
Table 2.A.2. Fixed effects summary for Experiment 1 in the Intervention conditions.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.245</td>
<td>0.202</td>
<td>16.050</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Wh1</td>
<td>1.041</td>
<td>0.105</td>
<td>9.900</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Wh2</td>
<td>0.122</td>
<td>0.105</td>
<td>1.159</td>
<td>.247</td>
</tr>
<tr>
<td>Wh1*Wh2</td>
<td>0.343</td>
<td>0.210</td>
<td>1.635</td>
<td>.103</td>
</tr>
</tbody>
</table>

Results showed a main effect of Wh1, attesting to a significant improvement in acceptability scores when Wh1 is lexically restricted as compared to when it is bare, showing that Complex Identity and Inclusion are globally rated higher than Bare Identity and Inverse Inclusion together. No main effect of Wh2 and no interaction between Wh1 and Wh2 were found. Subsequent models were conducted in order to provide a more fine-grained investigation of the relative acceptability of the 4 structures in the Intervention condition. A significant effect of Wh2 was attested when Wh1 is restricted ($\beta = 0.294, t = 1.898, p = .058$), showing that Complex Identity is rated higher than Inclusion, while no effect of Wh2 is attested when Wh1 is bare ($\beta = -0.050, t = -0.358, p = .721$), showing that Bare Identity and Inverse Inclusion are on par. In sum, results from individual models show the following pattern (where “>” means “more acceptable than” and “=” means “on a par with”): Complex Identity > Inclusion > Bare Identity = Inverse Inclusion.

No Intervention conditions. A summary of the fixed effects is reported in Table 2.A.3.

Table 2.A.3. Fixed effects summary for Experiment 1 in the No Intervention condition.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>5.916</td>
<td>0.140</td>
<td>42.098</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Wh1</td>
<td>-0.550</td>
<td>0.086</td>
<td>-6.397</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Wh2</td>
<td>-0.256</td>
<td>0.086</td>
<td>-2.980</td>
<td>.002</td>
</tr>
<tr>
<td>Wh1*Wh2</td>
<td>-0.650</td>
<td>0.172</td>
<td>-3.780</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Results reveal a main effect of Wh1, attesting to a significant improvement in acceptability scores when Wh1 is bare as compared to when it is lexically restricted, showing that Bare
Identity and Inverse Inclusion are globally rated higher than Inclusion and Complex Identity. A main effect of Wh2 was also found, attesting to higher acceptability ratings when Wh2 is bare than when it is lexically restricted, showing that Bare Identity and Inclusion are globally rated higher than Inverse Inclusion and Complex Identity. A significant interaction between Wh1 and Wh2 was also found, with a significant effect of Wh2 when Wh1 is restricted ($\beta = -0.581$, $t = -4.475$, $p < .001$), attesting to higher acceptability for Inclusion with respect to Complex Identity, but no effect of Wh2 when Wh1 is bare ($\beta = 0.068$, $t = 0.668$, $p = .504$), showing that no significant difference is attested between Bare Identity and Inverse Inclusion. In sum, results from individual models show the following pattern (where “>” means “more acceptable than” and “=” means “on a par with”): Bare Identity = Inverse Inclusion > Inclusion > Complex Identity.
Chapter 3

Similarity effects in wh-islands acceptability: A taxonomy of features

In six acceptability judgment experiments, I explored the role of similarity in wh-islands environments and in minimally different grammatical structures (that-clauses). Predictions stemming from two prominent theories accounting for similarity-based effects were systematically tested: Featural Relativized Minimality, a theory of grammar, and the Cue-based memory model, a theory of processing. Both theories assume that sentence acceptability is penalized by the similarity between the target element and the intervening element, but the former assumes that only morphosyntactic features triggering movement enter into the calculation of similarity (narrow similarity), while the latter assumes that any relevant linguistic feature does (broad similarity). I show that, in line with the processing theory, various types of features (lexical restriction, animacy and feature sets involved in thematic role reversibility) modulate the acceptability of both wh-islands and that-clauses. However, in accord with the grammatical theory, similarity in terms of these features has a weak effect compared to that of the [+Q] feature triggering movement associated with question operators. Capitalizing on insights from both theories, I suggest that two families of similarity effects have to be distinguished: effects due to identity in syntactic features triggering movement, defining grammatical violations, and effects due to identity in other linguistic features, in virtue of their role on memory encoding and retrieval mechanisms at play in the processing of sentences involving movement.

3.1 Introduction

In Chapters 1 and 2, I have discussed that according to Featural Relativized Minimality the catalogue of features that are relevant in the computation of locality are morphosyntactic
features triggering movement (Rizzi 2001, 2004). Moreover, I also discussed that among the features triggering movement, two families of features need to be distinguished: *criterial feature* (or *Type I features*), i.e., features that are able to trigger movement autonomously as [+Q], [+Rel], [+Top] and [+Foc], and *non-criterial features* (or *Type II features*), i.e., features that contribute in the fine identification of the landing site of movement only when accompanied by a Type I feature, as it is the case for the [+N] feature associated with lexical restriction. Beyond these two classes of features, no other feature is supposed to play a role in the computation of locality under fRM. I referred to these other features not triggering movement under any condition as Type III features. Featural Relativized Minimality thus endorses an approach to similarity that I characterize as *narrow similarity* in which only a subset of features, those triggering movement (i.e., Type I and II features), count in the calculation of similarity.

Empirical evidence in support of a narrow approach to similarity comes from a study by Belletti, Friedmann, Brunato and Rizzi (2012). In a picture matching task, the authors observed that the comprehension of object relative clauses improves when the object and the subject mismatch in gender in Hebrew speaking children aged 3;9-5;5 (e.g., *Show me the girl-FEM that the doctor-MASC draws-MASC* vs. *Show me the girl-FEM that the woman-FEM draws-FEM*), while no effect was found in Italian children. The authors attributed the discrepancy between Italian and Hebrew’s results to the fact that in Italian the verb does not agree in gender with its subject, unlike in Hebrew, and it therefore does not trigger movement in that language, while it does in Hebrew. The difference in the ability to trigger movement of the gender feature in the two languages has therefore be taken as an evidence that only features triggering movement can generate fRM-type effects.

This standpoint stands in contrast with the one endorsed by the Cue-based memory model (e.g., McElree 2000; Lewis and Vasishth 2005; Van Dyke and McElree 2006; McElree 2006). According to this model, the class of features playing a role in the definition of similarity extends beyond the features that trigger movement. In particular, any feature that reduces the similarity between the target and the distractor, either at encoding or at retrieval, enters in the computation of similarity. In this Chapter, I will refer to processing models that attribute a critical role to similarity-based interference in modulating processing difficulty as *Similarity-based Interference Memory Model* (henceforth, SIMM), in order to remain neutral about any
claim relative to the localization of the interference effect (retrieval vs. encoding), a question that I will address in Chapters 4 and 5.

Although SIMM endorses a broad approach to similarity, unlike fRM, its notion of similarity is not unconstrained. In particular, the resolution of long-distance dependencies is supposed to be conducted through a grammar-driven retrieval mechanism which is sensitive exclusively to those features that are linguistically relevant, i.e., that are likely to drive the parser during structure building, such as syntactic and semantic features. In line with this assumption, superficial features like the text color of to-be-retrieved element or its phonological similarity with other items in the sentence (i.e., the fact that the target may rhyme with other elements in the sentence or with elements in a memory load) has been shown to have no role whatsoever in the resolution of long-distance dependencies (see Hofmeister and Vasishth, 2014; Kush, Johns and Van Dyke 2015), even though these features play a role during earlier stage of processing, typically at encoding, a point to which I will return in Chapter 5. Although the SIMM model was initially developed to account for the difficulties observed in the comprehension of fully grammatical structures (like object relatives, clefts, or center-embedding structures), similarity-based interference has also been identified among the factors responsible for the poor acceptability of wh-islands (e.g., McElree et al. 2003; see Hofmeister and Sag 2010 for a discussion). However, a major limitation of this model, at least in its current formulation, is that it is not clear how it accounts for the fact that some long-distance dependencies, like wh-islands, are banned by the grammar, while some other long-distance dependencies, like object relatives, are difficult to process but ultimately grammatical.

3.2 Present study

The main hypothesis that I will defend in this Chapter capitalizes on insights from the two frameworks, fRM and SIMM, to provide what appears a better fit of the current empirical evidence at hand. I propose that similarity plays a key role in defining constraints both at the grammatical level and at the processing level. Two key sets of empirical evidence motivate this approach. First, the systematic investigation provided in Chapter 2 of structures involving long-distance dependencies and typically considered as ungrammatical (wh-islands and superiority violation) has revealed that their acceptability is improved when the two wh-elements are
lexically restricted (i.e., Complex Identity; e.g., *Which building do you wonder which engineer built?* or *Which building did which engineer build?*) compared to when only the moved element is lexically restricted and the intervening one is bare (i.e., Inclusion; e.g., *Which building do you wonder who built?* or *Which building did who build?*; see also Atkinson, Apple, Rawlins and Omaki 2015 for English findings and Hofmeister et al. 2007, 2013 for similar findings in Superiority violations in English). In Chapter 2, I adopted Rizzi’s proposal (2011) that Complex Identity may reduce to a configuration of Inclusion under the assumption that the extracted element targets a [+Q,+N] position, while the intervening element targets a [+Q] position, the parser being opportunistic in its attempt to maximize well-formedness. However, this still leaves unexplained the observation that Complex Identity is actually rated higher than the classical case of Inclusion (i.e., +Q,+N…+Q). As discussed in Chapter 2, the SIMM model can account for this finding by arguing that lexically restricted elements are endowed with a richer set of semantic features, making the two wh-elements more distinct and therefore easier to encode and/or retrieve from memory (Hofmeister et al. 2013).

Second, a recent study in English has shown that structures containing an extraction out of a wh-island and involving a configuration of Inclusion (i.e., +Q,+N…+Q, e.g., *Which of the cars do you wonder who might buy?*) are much less acceptable than structures that also involve a configuration of Inclusion but in which the extraction is out of a that-clause (i.e., +Q,+N…+N, e.g., *Which of the cars do you think that he might buy?*, Goodall 2015). This observation is in line with the one discussed in Chapter 1 (section 1.2.3) with respect to the case of object relatives, which also instantiate a configuration of inclusion (+Rel, +N…+N) but that are nonetheless perceived as much less degraded than wh-islands also instantiating a configuration of inclusion (+Q, +N…+Q). In order to account for this difference, I suggested to capitalize on Rizzi’s (1997, 2004) distinction between criterial features (i.e., Type I features), defined as features able to trigger movement independently of others (e.g., +Q, +R(el), +TOP, +Foc) and non-criterial features (i.e., Type II features), defined as features able to trigger movement only if accompanied by a criterial feature (e.g., +N). Following Rizzi (1997, 2004), I proposed that criterial features have a stronger role to play than non-criterial ones in the generation of intervention effects, such that an identity in criterial features with dissimilarity in non-criterial features (+Q,+N…+Q; criterial inclusion) is worse than an identity in non-criterial feature with dissimilarity in criterial features (+Q,+N…+N; non-criterial inclusion). As a first approximation, I suggested that an overlap on criterial features, as it is the case for wh-islands,
determines ungrammaticality, while an overlap on non-criterial features, as it is the case for object relatives or that-clause extraction, may determine processing difficulties without undermining the grammaticality of the sentence. However, I left aside the precise characterization of those structures that do involve an overlap on criterial features, but nonetheless satisfies the fRM principle, as Criterial Inclusion and Complex Identity, given that fRM is satisfied when the overlap in morphosyntactic features between the extractee and the intervener is not complete. Once that these structures are taken into account, the proposal sketched above amounts to assuming that two structures that both satisfy the fRM principle, namely Criterial and Non-criterial Inclusion, radically differ in their acceptability rates (Criterial Inclusion being much worse than Non-criterial Inclusion, see Goodall 2015), while two structures that are supposed one to violate fRM and the other to satisfy it, namely Bare Identity and Criterial Inclusion, only mildly differ in their acceptability rates (less than 1 point difference on a 7 point Likert scale based on the experiments reported in Chapter 2). In Chapter 2, I suggested to distinguish between three values of grammaticality: strong violation, corresponding to a full overlap of criterial features, weak violation, corresponding to a partial overlap of criterial features, and well-formedness, corresponding to a zero overlap of criterial features. This tripartite system indeed accounts for the observation that Criterial and Non-criterial inclusion, even though both satisfy the principle, are not on a par. However, this tripartite system still does not account for the improvement observed for Complex Identity as compared to Inclusion in the three experiments reported in Chapter 2. To account for this observation, one possibility, already addressed in Chapter 2 (section 2.6.1), would consist in amending the fRM system by enlarging the set of features entering in the calculation of the similarity, such that the lexical features expressed by the lexical restriction would also enter in the calculation of locality. However, as observed in Chapter 2, this revision would have undesirable consequences, as discussed by Belletti et al. 2012. A second possibility would consist in arguing that while the contrast between Bare Identity and Criterial Inclusion stems from an fRM-type effect, the contrast between Criterial Inclusion and Complex Identity is external to fRM, an option that has been discussed in Chapter 2, where this latter contrast has been accounted for in terms of memory mechanisms (section 2.6.2). This solution, although logically possible, does not seem to be very parsimonious as it would result in suggesting that a very same factor, namely lexical restriction, sometimes has an effect attributable to fRM and some other times has an effect not attributable to fRM. However, this possibility may obtain
additional plausibility if the effect size of lexical restriction was found to be consistently stronger in modulating the improvement of Criterial Inclusion as compared to Bare Identity than in modulating the improvement of Complex Identity as compared to Inclusion. I therefore estimated the effect size of lexical restriction in both contrasts in each of the three experiments presented in Chapter 2. Effect sizes were calculated through the Cohen’s d, which is obtained by dividing the mean difference between the two conditions by the pooled standard deviation of the two conditions. A Cohen’s d of 0.2 is usually considered as a small effect, a Cohen’s d of 0.5 as a medium effect and a Cohen’s d of 0.8 as a large effect (Cohen, 1988). Experiment 1 reports a medium effect of lexical restriction for the contrast between Bare Identity and Inclusion (Cohen’s d = 0.5) and a small effect for the contrast between Complex Identity and Inclusion (Cohen’s d = 0.2), Experiment 2 reports a tendency in the opposite direction, with the contrast between Complex Identity and Inclusion being bigger than the contrast between Bare Identity and Inclusion (Cohen’s d = 0.5 vs Cohen’s d = 0.4), and Experiment 3 reports comparable effect sizes for both contrasts (Cohen’s d = 0.3). All in all, results suggest that the effect of lexical restriction is essentially comparable in the two contrasts, even if some small fluctuation is observed, a finding that casts some doubts on the hypothesis that two different factors may be responsible for the improvement of lexical restriction in the two contrasts.

In this Chapter, I will explore a third possibility. I will suggest to distinguish between two families of similarity effects: effects due to similarity in terms of syntactic features triggering movement (i.e., Type I features), which play a role in defining the boundaries of grammar, and effects due to similarity in terms of other linguistic features (syntactic, semantic and possibly phonological, see Acheson and MacDonald 2011, and Kush, Johns and Van Dyke 2015, i.e., Type II and III features), which play a role in the processes of memory retrieval and encoding involved in the parsing of long-distance dependencies. In line with early formulations of fRM, the presence of an intervening element endowed with features triggering movement identical to those of the moved element blocks the syntactic dependency between the moved element and its trace (Rizzi 2001, 2004). I suggest to restrict the set of features triggering movement to those referred to as criterial by Rizzi (2004) (i.e., +Q, +R(el), +TOP, +Foc), and consider lexical restriction as a morphosyntactic feature with semantic correlates, whose effect is thus expected to be on a par with that of other linguistic features not triggering movement. In line with SIMM, I argue that the process of retrieving the moved element from memory is penalized by the presence of other similar elements in the sentence, either because similarity penalizes
encoding (through feature overwriting) or because it penalizes retrieval (through the cue-based retrieval process sensitive to cue-overlad). This hypothesis makes the following predictions, tested in a series of 6 experiments conducted in French:

(a) Similarity in terms of features that do not trigger movement should affect the acceptability of wh-islands. Experiments 1 and 2 test, by way of an acceptability judgment procedure, the influence of animacy, a syntactic-semantic feature, and of the reversibility of thematic roles, a set of semantic features. Both Experiments 1 and 2 also test the role of lexical restriction, a syntactic feature with a semantic correlate. Featural RM predicts no effect of animacy and reversibility, since neither of them triggers movement. Under the assumption that lexical restriction triggers movement of wh-elements, an effect of that feature is expected in wh-islands. SIMM predicts a significant effect of all three features.

(b) These three features should also have an effect in structures involving long-distance dependencies and typically considered as grammatical. Experiments 3 and 4 test the effect of animacy and reversibility in both wh-islands and that-clauses, with a two-alternative forced choice task, which was shown to be more powerful than the Likert scale (Sprouse and Almeida, 2017). In contrast to Goodall (2015), who claimed that fRM is exclusively concerned with predictions about wh-islands while no prediction can be made for that-clause extraction, I argue that fRM does make predictions about that-clauses, because, as already discussed in Chapter 1, fRM applies to any sentence involving a configuration of intervention, be grammatical or not (see Friedmann et al. 2009, Belletti et al. 2012). Following the same logic, fRM predicts that that-clauses, like wh-islands, will be insensitive to animacy or reversibility, since none of these

---

1 In French, animacy is a morphological feature realized on the interrogative pronoun (qu’ the contracted form of que (what) for inanimate referents and qui (who) for animate referents). Although there are languages for which it is arguable that animacy is a feature triggering movement (this could be the case for differential object marking in Spanish, see for instance Torrego 1998, for Romanian, see Farkas 1978), this does not seem to be the case for French given that this feature does not belong to the phi-feature set expressed in the verbal morphology (see Belletti et al. 2012, Bentea and Durrlemann 2014). In what follows, I therefore assume that animacy does not trigger movement in French in line with previous works making the same assumption (e.g., Bentea and Durrlemann 2014; but see Bentea et al. 2016 for a different standpoint).
features triggers movement. On the other side, SIMM predicts an effect of both these features in both wh-islands and that-clauses, since both types of structures involve long-distance dependencies and thus memory encoding/retrieval processes sensitive to similarity.

(c) Similarity in terms of features triggering movement should have a strong effect on sentence acceptability. Experiment 5 tests the effect of the +Q feature defining question operators and triggering movement with a two-alternative forced choice task, while Experiment 6 provides a direct comparison of the effect size of the +Q feature and the effect size of lexical restriction (+N) in an acceptability judgment task, replicating the English experiment of Goodall (2015). Whereas fRM predicts that +Q will have a stronger effect than +N, the former being a criterial feature, SIMM, at least in its current formulation, predicts no difference.

3.3 Experiment 1: The effect of animacy and lexical restriction in wh-islands

3.3.1 Method

Participants. Forty-two French-speaking students at the University of Geneva took part in the experiment for course credit. They were between 18 and 26 years of age and were all native French speakers.

Material. Two variables were manipulated: i) the Lexical Restriction of the wh-elements (both bare vs. both lexically restricted), and ii) the match in Animacy between the extracted wh-element and the intervening wh-element (animacy match, where both wh-elements are animate vs. animacy mismatch, where the extracted wh-element is inanimate and the intervening wh-element is animate). The two variables were part of a fully-crossed design involving 4 experimental conditions. All verbs required animate subjects, but they could either take animate or inanimate objects. Half of the experimental sentences contained *se demander* (wonder) as a main verb, whereas the other half contained *savoir* (know). All questions containing the verb *se demander* were affirmative, while half of the questions containing the verb *savoir* were negative, with the aim of introducing some variability in the materials.
Thirty-two experimental items were generated. Each of them appeared in the 4 experimental conditions such that a total of 128 experimental sentences were created, and split into two between-subject lists containing 64 items each. Therefore, each subject read two conditions per item (either the animacy match or the animacy mismatch conditions), and read an equal number of match and mismatch sentences. The experimental sentences of each list were intermixed with 136 fillers, identical across lists. Fillers consisted of wh-islands, including the experimental items of Experiment 2.

**Procedure.** Each sentence was displayed on a computer screen one at a time. Participants were asked to carefully read each sentence and then to judge its acceptability on a 7-point Likert-scale (1 corresponding to a wholly unacceptable sentence and 7 to a fully acceptable one) by pressing one of the seven numbered buttons on the keyboard. We explicitly informed participants that there was no time constraint on responses. Participants were first presented with 10 training items in order to familiarize them with the Likert scale. Three short pauses of 1 minute each were administered during the task. The whole session lasted about 20 min.

---

2 Notice that the term *bare* refers to the Bare Identity condition in which both wh-elements are bare, and the term *restricted* refers to the Complex Identity condition, in which both wh-elements are restricted.
3.3.2 Results

**Data analysis and results.** Figure 3.1 shows the mean acceptability in z-scores for the four conditions of Experiment 1. Zero represents the grand mean. Hence, a positive z-score means that the structure is rated above the grand mean, while a negative z-score means that the structure is rated below the grand mean. In addition, the z-score tells us how many standard deviations below or above the mean a raw score is (e.g., a z-score equal to 1 refers to a structure that is 1 standard deviation greater than the grand mean). Raw scores are reported in the Appendix at the end of this Chapter. A linear mixed-effects model was fitted to the data using the lmerTest package (http://www.cran.r-project.org/web/packages/lmerTest//lmerTest.pdf) in the R software environment (R Core Team, 2016), with random intercepts and slopes for both subjects and items, and lexical restriction and animacy as predictors (fixed factors). Analyses are therefore conservative with respect to the generalizability of the effects of theoretical interest to new participants and items (Barr, Levy, Cheepers, and Tily, 2013). All predictive factors were dichotomous and centred by coding one level of the factor as -1 and the other as 1. P-values were calculated by way of the Satterthwaite’s approximation to degrees of freedom with the lmerTest package (Kuznetsova et al. 2016). Prior to analysis, acceptability judgments from each participant were transformed to z-scores in order to correct for scale bias, but the same results were obtained on non-transformed scores. A summary of the fixed effects is reported in Table 3.1.
Figure 3.1. Mean acceptability (z-scores) for the 4 conditions of Experiment 1. Error bars indicate standard errors.

Analyses revealed a main effect of lexical restriction, attesting to significantly higher ratings for sentences with restricted wh-elements than for sentences with bare wh-elements ($p < .001$). No main effect of animacy was observed ($t < 2$), but animacy entered into a significant interaction with lexical restriction ($p < .001$). Separate models were conducted to assess the role of animacy for bare and restricted conditions. In the restricted condition, higher ratings were found for sentences with animacy mismatch than for sentences with animacy match ($\beta = -0.139$, $t = 5.714$, $p < .001$). No effect of animacy was found in the bare condition ($t < 2$).

Table 3.1. Summary of the fixed effects for Experiment 1.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.181</td>
<td>0.029</td>
<td>-6.168</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Lexical Restriction</td>
<td>0.399</td>
<td>0.048</td>
<td>8.174</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Animacy</td>
<td>0.038</td>
<td>0.024</td>
<td>1.618</td>
<td>.113</td>
</tr>
<tr>
<td>Lexical Restriction*Animacy</td>
<td>0.199</td>
<td>0.026</td>
<td>3.861</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

3.3.3 Discussion

Results from Experiment 1 show that conditions with two restricted wh-elements are more acceptable than those with two bare wh-elements. This result replicates findings from Chapter 2 on wh-islands (see also Atkinson et al. 2015 for English) as well as previous findings on superiority violations (Hofmeister et al. 2013). This result is in line with the prediction of fRM, if one assumes not only that lexical restriction triggers movement in French, but also that the parser has the option to opportunistically treat the restricted condition as a configuration of feature inclusion (Rizzi 2011). However, in contrast to fRM’s predictions, results also attest to a significant improvement of wh-islands’ acceptability when the two elements mismatch in animacy, if the two wh-elements are lexically restricted. Both these findings find a natural explanation under the broad similarity approach endorsed by SIMM, according to which
sentences containing syntactically and semantically more distinctive elements are easier to parse. The semantic richness of restricted wh-elements like *which class* and *which student* makes them more distinct from one another than two bare wh-elements like *what* and *who*, thus reducing similarity-based interference and, as a result, increasing sentence acceptability (see Hofmeister et al. 2007, 2013).

No effect of animacy was observed in sentences with two bare wh-elements, a result which aligns with findings in French children showing that although animacy affects the comprehension of wh-object questions when the wh-object is lexically restricted (e.g., *Which ball is the girl hitting?* is better understood than *Which lady is the girl kissing?*), it does not have an effect when the wh-object is bare (e.g., *What is the girl hitting?* is on a par with *Who is the girl hitting?*) (Bentea, Durrleman and Rizzi 2016). To account for the lack of an animacy effect in the bare condition, the authors hypothesized that the way in which animacy is grammatically encoded matters in the calculation of similarity, such that only when the feature is uniformly expressed on the target and on the intervener the system takes it into account. This means that when the feature is expressed either on the wh-determiner or on the lexical restriction in both the target and the intervener, the system takes it into account, while it does not otherwise. However, this hypothesis does not account for the current results, since even if the animacy feature is expressed uniformly in the bare condition, no effect is observed. This finding may be due to the fact that even though animacy, as a syntactic feature, is present on both nouns and question operators, animacy is also a semantic feature, which may be more salient as a property of nouns than as a property of functional particles. Another possibility is that the lack of effect in the bare conditions is due to a floor effect, although this would not explain the results obtained by Bentea et al. (2016). In either case, the effect of animacy in restricted conditions is enough to conclude that similarity in animacy significantly modulates acceptability ratings in wh-island environments, contra the narrow approach of similarity endorsed by fRM.³

---
³ Bentea, Durrleman and Rizzi (2016), capitalizing on Bianchi’s (2006) Person-Case constraint in Romance languages, suggested that animacy may trigger movement in French (and possibly universally), in which case its effect would be accounted for by fRM.
In this experiment, conditions with a mismatch in animacy always contained sentences with non-reversible thematic roles while sentences with a match in animacy were always reversible. One thus cannot exclude the possibility that reversibility, rather than animacy, is responsible for the ameliorative effect found for sentences with an animacy mismatch. With the aim of disentangling these two factors, I ran Experiment 2, manipulating the reversibility of thematic roles, keeping animacy constant. Since reversibility depends solely on the semantics of the noun phrases and their relation to the verb, any effect of this variable on wh-island acceptability would provide strong support for the broad similarity hypothesis.

3.4 Experiment 2: The effect of reversibility and lexical restriction in wh-islands

3.4.1 Method

Participants. The same participants recruited for Experiment 1 took part in Experiment 2.

Material. Thirty-two experimental items were generated. Two variables were manipulated: i) the Lexical Restriction of the wh-elements (both bare vs. both lexically restricted), and ii) the Reversibility of thematic roles (reversible vs. non-reversible). Reversible sentences contained arguments that could both plausibly be the agent or the patient of the verb (e.g., a boss seeing an employee is as plausible as an employee seeing a boss), while non-reversible sentences contained arguments whose thematic roles were not interchangeable (e.g., a boss may fire an employee but an employee may not fire a boss). The two variables were part of a fully-crossed design involving 4 experimental conditions. All wh-elements were animate. The choice of the main verb (i.e., wonder, know) and its polarity was the same as in Experiment 1.

(5) Qui te demandes-tu qui a licencié? (Bare, Non-reversible)
    ‘Who do you wonder who fired?’

(6) Qui te demandes-tu qui a vu? (Bare, Reversible)
    ‘Who do you wonder who saw?’

(7) Quel employé te demandes-tu quel chef a licencié? (Restricted, Non-reversible)
    ‘Which employee do you wonder which boss fired?’
(8)  Quel employé te demandes-tu quel chef a vu?  (Restricted, Reversible)

‘Which employee do you wonder which boss saw?’

The experimental items were split in two between-subject lists containing 64 items each and intermixed with 136 fillers, all constituted by wh-islands, amongst which the experimental items of Experiment 1.

Procedure. The procedure was the same as in Experiment 1.

3.4.2 Results

Data analyses and results. The same data analyses as in Experiment 1 were conducted. Figure 3.2 shows the mean acceptability in z-scores for the four conditions of Experiment 2, while a summary of the fixed effects is reported in Table 3.2. Analyses on the raw scores are reported in Appendix at the end of the Chapter.

![Figure 3.2. Mean acceptability in z-scores for the 4 conditions of Experiment 2. Error bars indicate standard errors.](image)

Results attested to a significant main effect of lexical restriction ($p < .001$), with higher scores when both wh-elements are restricted than when they are bare. Moreover, a main effect of reversibility was also observed ($p = .001$), attesting to higher scores when the sentence is non-
reversible than when it is reversible. Finally, a significant interaction between lexical restriction and reversibility was also observed ($p = .005$). Separate models were conducted to assess the role of reversibility for bare and restricted conditions. In the restricted condition, higher ratings were found for the non-reversible condition than for the reversible one ($\beta = 0.099, t = 4.032, p < .001$). No effect of reversibility was found in the bare condition ($t < 1$).

Table 3.2. Summary of the fixed effects for Experiment 2.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.197</td>
<td>0.032</td>
<td>-6.170</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Lexical Restriction</td>
<td>0.299</td>
<td>0.047</td>
<td>6.283</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Reversibility</td>
<td>0.053</td>
<td>0.016</td>
<td>3.299</td>
<td>.001</td>
</tr>
<tr>
<td>Lexical Restriction* Reversibility</td>
<td>0.046</td>
<td>0.016</td>
<td>2.845</td>
<td>.005</td>
</tr>
</tbody>
</table>

3.4.3 Discussion

The results replicated the finding of Experiment 1 that sentences with two restricted wh-elements are rated higher than those with two bare wh-elements. In addition, the results showed a significant effect of reversibility, with higher acceptability ratings for non-reversible sentences than for reversible ones. That is, sentences in which the object was not a plausible agent of the action were rated higher than those in which the object was a plausible agent. This factor only played a role in sentences containing restricted wh-elements, suggesting that what matters is the semantics of the verb-arguments relationship, and not the specific selectional constraints of the verbs used in the two conditions. The finding that reversibility modulates acceptability judgments is in line with the broad approach to similarity of the SIMM, but is not accounted for by fRM.

Given that reversibility turns out to modulate acceptability ratings, the question arises whether the lower ratings obtained for restricted conditions with a match in animacy in Experiment 1 (in which thematic roles were reversible) as compared to conditions with a mismatch in animacy (in which thematic roles were non-reversible) were entirely due to reversibility, or whether animacy still played a role on top of the reversibility effect. With the aim of evaluating the relative roles of animacy and reversibility, I compared results from the animacy mismatch
condition with non-reversible thematic roles of Experiment 1 with the animacy match condition with non-reversible thematic roles of Experiment 2. If animacy plays a role in addition to reversibility, we expected to find a significant improvement in the animacy mismatch (non-reversible) condition of Experiment 1 as compared to the animacy match (non-reversible) condition of Experiment 2.

3.5 Merge of Experiments 1 and 2

I report results from the comparison between the condition with animacy mismatch (non-reversible) of Experiment 1 and the condition with animacy match (non-reversible) of Experiment 2, the two conditions involving lexically restricted wh-elements. Crucially, while in the animacy mismatch condition the reversibility is prevented by animacy itself, in the animacy match condition the reversibility is prevented by the semantics of the verb-argument relationship. I ran a mixed-effects model with animacy as fixed factor. A summary of the fixed effects is reported in Table 3.3. Results revealed a significant effect of animacy ($p = 0.048$), attesting to higher ratings in the animacy mismatch condition than in the match condition.

Table 3.3. Summary of the fixed effects for the Experiment 1 and Experiment 2 merged together.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.201</td>
<td>0.062</td>
<td>3.248</td>
<td>.001</td>
</tr>
<tr>
<td>Animacy</td>
<td>0.155</td>
<td>0.077</td>
<td>2.018</td>
<td>.048</td>
</tr>
</tbody>
</table>

In sum, both animacy and reversibility participate in modulating acceptability ratings, although animacy appears to play a greater role than reversibility since it has an effect on non-reversible sentences as well. Hence, the highest rated condition is the one in which the wh-elements mismatch in animacy and they are non-reversible, followed by the condition in which the wh-elements match in animacy and they are non-reversible, while the lowest rated condition is the one in which both wh-elements are animate and the thematic roles are reversible.

Do animacy and reversibility specifically modulate acceptability ratings in wh-island configurations or does their effect extend to other structures? There is evidence in the literature showing that animacy and reversibility play a role in the comprehension of fully grammatical
structures, in particular object relatives. For example, the well-known advantage of subject relative clauses over object relatives can be neutralized when the head of the object relative is inanimate and the subject is animate (see Mak et al. 2002, 2006, Traxler et al. 2002, Gennari and MacDonald 2008 for adults; Corrêa 1995, Arosio et al. 2010, Bentea and Durrleman 2014, Bentea, Durrleman, Rizzi 2016 for children). Moreover, sentences containing a subject and an object with non-reversible thematic roles are easier to comprehend than those with reversible roles (see, among others, Caramazza and Zurif 1976 in typical and atypical adults, van der Lely and Harris 1990, Stavrakaki 2001 in specifically language impaired children, and Garraffa and Grillo 2008 in aphasics). In order to finely assess the role of animacy and reversibility in wh-islands and in minimally different structures considered as grammatical, Experiments 3 and 4 investigated their effect in both wh-islands and that-clause extractions. I restricted the investigation to lexically restricted sentences, which were the locus of the effect of these two variables in wh-islands. A two-alternative forced choice task was used, in which we asked participants to select the more acceptable of two sentences differing on the variable manipulated. Forced choice has been shown to be more powerful than other acceptability methods (Sprouse and Almeida, 2017), especially to detect fine acceptability differences, and therefore a promising way to capture animacy and reversibility effects given their mild effect in modulating acceptability ratings found in Experiments 1 and 2.

3.6 Experiment 3: The effect of animacy in wh-islands and that-clauses

3.6.1 Method

Participants. Sixty French-native speaker participants took part in the experiment (mean age = 29 y.o.). They were uniformly distributed across lists. Participants were naïve to the purpose of the experiment and they did not take part neither in Experiment 1 nor in Experiment 2.

Material. Twelve items were tested, which were a subset of the materials used in Experiment 1, but in addition to Experiment 1, a that-clause condition was added. Participants were asked to directly compare sentences with a mismatch in animacy to sentences with a match in animacy.
both in wh-islands (9a vs. 9b) and in that-clause environments (10a vs. 10b) and select the most acceptable sentence of the pair. The resulting 24 contrasts (12 contrasts testing for animacy in wh-islands and 12 constrasts manipulating for animacy in that-clauses) were split into 6 lists. Each list contained only 4 contrasts, 2 testing for animacy in wh-islands and 2 testing for animacy in that-clauses, thus reducing the chances of satiation effects (i.e., the increase in acceptability after repeated exposures, Snyder 2000). Sprouse and Almeida (2017) showed that forced choice experiments reach 80% power with one observation per participant per condition and 30 participants. Questions were always formed through subject-verb inversion. Only lexically restricted elements were used.

(9)  
a. Quel cours te demandes-tu quel étudiant a apprécié? (Wh-island, mismatch)  
‘Which class do you wonder which student appreciated?’

b. Quel professeur te demandes-tu quel étudiant a apprécié? (Wh-island, match)  
‘Which professor do you wonder which student appreciated?’

(10)  
a. Quel cours crois-tu que l’étudiant a apprécié? (That-clause, mismatch)  
‘Which class do you believe that the student appreciated?’

b. Quel professeur crois-tu que l’étudiant a apprécié? (That-clause, match)  
‘Which professor do you believe that the student appreciated?’

**Procedure.** The experiment was programmed using Qualtrics ([http://www.qualtrics.com](http://www.qualtrics.com)), an online experimental platform. The sentences, arranged in pairs, were presented one above the other and participants were asked to select the sentence that they found more acceptable. Pairs were always lexically matched to form minimal contrasts that only diverged on the syntactic property of interest, i.e., animacy. Each pair was displayed on a computer screen one at a time. There was no time constraint. Items were paired such that participants were asked to select the most acceptable sentence between 9a and 9b and the most acceptable sentence between 10a and 10b. Hence, the effect of animacy was investigated separately for wh-islands and that-clauses. The whole session lasted about 5 minutes.

**3.6.2 Results**

**Data analysis and results.** Responses were coded as 1 when participants selected the condition with animacy mismatch and as 0 when they selected the condition with animacy match. I then calculated the proportions of responses, which therefore reflect the proportion of cases in which
the animacy mismatch condition was selected over the animacy match condition, respectively in wh-islands and that-clauses. I then conducted a logistic regression model using the lme4 package (Bates et al., 2015) in R (R Core Team, 2016) with sentence type (wh-islands vs. that-clause) as fixed factor. Contrasts were sum-coded (that-clauses were set as -1 and wh-islands as +1), such that the intercept of the model represents the grand mean response rate. Therefore, a positive, significant intercept means that the mismatch condition is significantly higher than the chance level, while the significance of the main effect will inform us whether animacy has the same effect in islands and that-clause.

A summary of the fixed effects is reported in Table 3.4. Results indicated that participants selected the animacy mismatch condition 64% of the times in the wh-island condition and 59% of the times in the that-clause condition. The positive, significant intercept ($p = .02$) indicates that both percentages are above the chance level. Moreover, the two percentages did not differ ($p = .248$) suggesting that the effect of animacy in wh-islands and in that-clauses was comparable.

**Table 3.4.** Summary of the fixed effects for the Experiment 3.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.469</td>
<td>0.205</td>
<td>2.289</td>
<td>.022</td>
</tr>
<tr>
<td>Animacy</td>
<td>0.201</td>
<td>0.174</td>
<td>1.154</td>
<td>.248</td>
</tr>
</tbody>
</table>

**Figure 3.3.** Proportion of animacy mismatch choices in that-clauses and wh-islands in Experiment 3. Error bars indicate standard errors. The horizontal dotted line indicates chance
3.6.3 Discussion

Experiment 3 replicates, with a new procedure, results from Experiment 1 showing that sentences with a mismatch in animacy are judged more acceptable than sentences with a match in animacy. Importantly, animacy has a similar effect in that-clause and wh-island conditions, suggesting that a mismatch in animacy ameliorates acceptability judgments in both wh-islands and that-clauses to a similar extent. This finding is in line with SIMM according to which memory processes underlying long-distance dependency processing are sensitive to similarity-based interference, whether the sentence is an island or not. Although that-clauses, like wh-islands, involve an intervention configuration in which the subject intervenes on the dependency between the moved object and its trace, fRM fails to account for the animacy effect in both of these structures, since animacy is not a feature triggering movement and is therefore assumed to be irrelevant to intervention effects. In Experiment 4 I investigated if the same conclusions can be reached for the reversibility of thematic roles.

3.7 Experiment 4: The effect of reversibility in wh-islands and that-clauses

3.7.1 Method

Participants. The same participants who took part in Experiment 3 also took part in this experiment.

Material. Twelve items of Experiment 2 were tested and thematic role reversibility (reversible vs. non-reversible) was manipulated in wh-islands and that-clauses. Participants were asked to
directly compare sentences with non-reversible thematic roles and sentences with reversible thematic roles by selecting the most acceptable one both in wh-islands (11a vs. 11b) and in that-clause environments (12a vs. 12b). Only sentences with lexically restricted elements were tested. The experimental items were split in 6 lists, as in the previous experiment.

(11) a. Quel employé te demandes-tu quel chef a licencié? (Wh-island, Non-reversible)  
‘Which employee do you wonder which boss fired?’

b. Quel employé te demandes-tu quel chef a vu? (Wh-island, Reversible)  
‘Which employee do you wonder which boss saw?’

(12) a. Quel employé crois-tu que le chef a licencié? (That-clause, Non-reversible)  
‘Which employee do you believe that the boss fired?’

b. Quel employé crois-tu que le chef a vu? (That-clause, Reversible)  
‘Which employee do you believe that the boss saw?’

Procedure. The same procedure as Experiment 3 was adopted.

3.7.2 Results

Data analyses and results. The same analyses were conducted as for Experiment 3. Responses were coded as 1 when participants selected the non-reversible condition and as 0 when they selected the reversible condition. A summary of the fixed effects is reported in Table 3.5. Participants selected the non-reversible condition in 70% of the cases in wh-island sentences and in 77.5% of the cases in that-clause sentences. The positive, significant intercept ($p < .001$) indicates that both percentages are above the chance level. Moreover, the two percentages did not differ ($p = .541$), suggesting that the effect of reversibility is similar in wh-islands and that-clause.

Table 3.5. Summary of the fixed effects for the Experiment 4.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.240</td>
<td>0.266</td>
<td>4.657</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Reversibility</td>
<td>-0.125</td>
<td>0.205</td>
<td>-0.611</td>
<td>.541</td>
</tr>
</tbody>
</table>
Figure 3.4. Proportion of selection of non-reversible sentences over reversible sentences in that-clauses and wh-islands in Experiment 4. Error bars indicate standard errors. The horizontal dotted line indicates chance level.

3.7.3 Discussion

Experiment 4 aligns with results from Experiment 2 in showing, with a new procedure, that non-reversible wh-islands are more acceptable than reversible ones. Moreover, the reversibility effect was also observed in that-clause sentences, to a similar extent as in wh-island sentences. This result is again in line with the SIMM since arguments with non-reversible roles are semantically less similar, and therefore less prone to similarity-based interference when the long-distance object needs to be retrieved from memory. The data cannot be accounted by fRM, since reversibility can not be analyzed as a feature triggering movement.

Altogether, Experiments 1-4 show that both animacy and thematic role reversibility modulate the acceptability of sentences involving a long-distance dependency between their object and the verb, whether the sentence is classically considered as ungrammatical (wh-islands) or whether it is fully grammatical (that-clauses). The next question I want to address is whether so-called criterial features, i.e., features triggering movement (like +Q, +Rel, +TOP, +Foc), play a special role in sentence acceptability, as assumed under fRM (Rizzi 2004). Experiments
5 and 6 explore the role of the overlap in terms of the +Q feature, by comparing wh-islands and that-clauses. Experiment 5 uses the same two-alternative forced-choice procedure as Experiments 3 and 4, while Experiment 6 explores its role with an acceptability judgment task like Experiments 1 and 2.

3.8 Experiment 5: The effect of criterial features

3.8.1 Method

Participants. The same participants that took part in Experiments 3 and 4 also participated in this Experiment.

Material. The sentences were the same as Experiments 3 and 4, but participants were asked to select the most acceptable sentence between a sentence with an island structure (13b and 14b) and the minimally different that-clause structure (13a and 14a). Half of the sentences had animacy mismatch and non-reversible thematic roles and the other half had animacy match and reversible thematic roles. The experimental items were split into 6 lists.

(13) a. Quel cours crois-tu que l’étudiant a apprécié? (That-clause, mismatch, non-reversible)
   ‘Which class do you believe that the student appreciated?’
 b. Quel cours te demandes-tu quel étudiant a apprécié? (Wh-island, mismatch, non-reversible)
   ‘Which class do you wonder which student appreciated?’

(14) a. Quel étudiant crois-tu que le professeur a vu? (That-clause, match, reversible)
   ‘Which student do you believe that the professor saw?’
 b. Quel étudiant te demandes-tu quel professeur a vu? (Wh-island, match, reversible)
   ‘Which student do you wonder which professor saw?’

Procedure. The same procedure used in Experiments 3 and 4 was used here.
3.8.2 Results

**Data analyses and results.** The same analyses were conducted as for Experiment 3. Responses were coded as 1 when participants selected the that-clause and as 0 when they selected the wh-island. A summary of the fixed effects is reported in Table 3.6. Participants selected the that-clause over the wh-island in 97.5% of the sentences that contained an animacy mismatch (non-reversible thematic roles) and in 96.6% of the sentences that contained an animacy match (reversible thematic roles) condition. The positive, significant intercept ($p < .001$) indicates that both percentages are above the chance level. Moreover, the two percentages did not differ ($p = .697$).

Table 3.6. Summary of the fixed effects for the Experiment 5.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>7.659</td>
<td>2.118</td>
<td>3.616</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Reversibility</td>
<td>0.2007</td>
<td>0.516</td>
<td>0.388</td>
<td>.697</td>
</tr>
</tbody>
</table>

![Figure 3.5](image_url) Proportion of selection of that-clause sentences over wh-islands in animacy mismatch, non reversible and animacy match, reversible conditions of Experiment 5. Error bars indicate standard errors. The horizontal dotted line indicates chance level.
3.8.3 Discussion

Results from Experiment 5 reveal that that-clauses are preferred to wh-islands in 97% of the cases (averaging between animacy mismatch, non-reversible and match, reversible sentences). In other words, wh-islands were virtually never selected, thus confirming that wh-islands are perceived as highly degraded as compared to that-clauses. Under the viewpoint of fRM, the clear dispreference for islands is due to the overlap in terms of a criterial feature triggering movement (+Q), which was argued to play a special role in intervention effects (Rizzi 2001, 2004). The results show a categorical shift between wh-islands and that-clauses, contrasting with the mild preference observed in relation to animacy and thematic role reversibility, which lie between 60% and 75%. However, preferences observed with a forced-choice procedure are difficult to compare, since the measure is categorical rather than graded: a preference for sentence A over sentence B does not tell us how much A is preferred. In order to be able to directly compare the effect size of an overlap in terms of the +Q criterial feature to that of an overlap in terms of lexical restriction, a non-criterial feature which does not trigger movement by itself, I conducted Experiment 6, which uses an acceptability judgment task.

3.9 Experiment 6: The effect of criterial features and lexical restriction

3.9.1 Method

Participants. Forty-nine French native speakers who did not take part in any of the other experiments participated in Experiment 6. Participants were asked to judge the acceptability of the sentences on a 7-point Likert scale.

Material. Eight items were generated by manipulating: (i) the type of structure (that-clause vs. wh-island), and (ii) the lexical restriction of the elements (both bare vs. both lexically restricted). In the bare, that-clause condition, the intervening subject was a pronoun (he), while it was a definite description (the student) in the restricted condition, under the hypothesis that
pronouns are less specified than definite descriptions, mirroring the distinction between bare and restricted wh-elements (Elbourne, 2005). All extracted objects were inanimate, while all subjects were animate. Half of the wh-islands contained *se demander* (wonder) as a main verb, while the other half contained *savoir* (know). All that-clauses contained the verb *croire* (believe). The experimental items were intermixed with 96 filler sentences consisting of wh-islands, superiority violations as well as fully grammatical wh-in situ questions, and then split into 4 lists.

(22) Qu’est-ce que tu crois qu’il a résolu?

‘What do you believe that he solved?’

+Q +Pro

(23) Quel problème crois-tu que l’étudiant a résolu?

‘Which problem do you believe that the student solved?’

+Q,+N +N

(24) Qu’est-ce que tu te demandes qui a résolu?

‘What do you wonder who solved?’

+Q,+N +Q

(25) Quel problème te demandes-tu quel étudiant a résolu?

‘Which problem do you wonder which student solved?’

+Q,+N +Q,+N

**Procedure.** The same procedure was used as for Experiments 1 and 2.

### 3.9.2 Results

**Data analyses and results.** The same data analyses as in Experiment 1 were conducted. In addition, I estimated the effect size of the variables through Cohen’s *d*, which is calculated by dividing the mean difference between the two conditions by the pooled standard deviation of the two conditions. A Cohen’s *d* of 0.2 is considered as small effect, a *d* of 0.5 a medium effect and a *d* of 0.8 a large effect (Cohen, 1988). Figure 3.6 shows the mean acceptability in *z*-scores for the four conditions of Experiment 6, while a summary of the fixed effects is reported in Table 3.7, and the raw scores are reported in Appendix at the end of this Chapter.
Figure 3.6. Mean acceptability in z-scores for the 4 conditions of Experiment 6. Error bars indicate standard errors. The dotted line indicates fRM’s grammaticality cut off point of grammaticality, while the dashed line indicates the cut off point if only criterial features play a role in fRM, as I will propose below (see General Discussion).

Results attested to a significant main effect of the type of structure ($p < .001$), with higher acceptability ratings for that-clauses than for wh-islands. A main effect of lexical restriction was also observed with higher scores when both elements are restricted than when they are bare. Finally, a significant interaction between the type of structure and lexical restriction was attested ($p = .001$). Additional models revealed a significant effect of lexical restriction in both that-clauses ($\beta = -0.061, \ SE = 0.028, \ t = -2.141, \ p = .034$) and wh-islands ($\beta = -0.289, \ SE = 0.066, \ t = -4.334, \ p = .003$). The effect size of the type of structure is well above the margin for a large effect (Cohen’s $d = 1.5$), while the effect size of lexical restriction is small (Cohen’s $d = 0.3$).
### Table 3.7. Summary of the fixed effects for Experiment 6.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.010</td>
<td>0.051</td>
<td>0.202</td>
<td>0.856</td>
</tr>
<tr>
<td>Type of structure</td>
<td>0.764</td>
<td>0.033</td>
<td>23.108</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Lexical Restriction</td>
<td>-0.180</td>
<td>0.031</td>
<td>-5.867</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Type of structure* Lexical Restriction</td>
<td>0.118</td>
<td>0.034</td>
<td>3.442</td>
<td>.001</td>
</tr>
</tbody>
</table>

#### 3.9.3 Discussion

In line with results from Goodall (2015) in English, Experiment 6 reveals that lexical restriction not only affects the acceptability of wh-islands but it also affects fully grammatical that-clause sentences in French. However, the estimate of the effect size of this variable shows that it is a small effect ($d = 0.3$), in contrasts to the very large size of the effect of structure ($d = 1.5$). The effect size of lexical restriction is comparable to that of animacy in Experiment 1 ($d = 0.24$) and to that of reversibility in Experiment 2 ($d = 0.1$).

In contrast to Goodall (2015), the effect of lexical restriction is stronger for wh-islands than for the corresponding grammatical sentences. However, a closer look at the English study shows that the interaction was actually close to being significant ($p = .062$). The two data sets differ in that the French ratings in the grammatical condition are nearly at ceiling, which is not the case in the English study; the interaction found in French may therefore be due to the fact that the margin for the effect of lexical restriction in the grammatical condition was simply too small (Loftus 1978). It is notable that, despite the near ceiling ratings, lexical restriction still showed a significant effect on the acceptability of that-clause sentences. The finding that lexical restriction affects that-clauses is in line with our finding that animacy and reversibility also influenced the acceptability of these fully grammatical structures in Experiments 3 and 4.

#### 3.10 General Discussion

In six experiments on the acceptability of French wh-islands and that-clauses, I tested the effect of similarity between a moved object and the intervening subject in terms of different types of features: lexical restriction, animacy, reversibility and the +Q feature. Both fRM and SIMM
grant a key role to the similarity between the moved element and the intervening one: the more their featural specifications overlap, the more the sentence is expected to be degraded. However, the two accounts differ on the catalogue of features that are supposed to enter into the similarity metric. According to the narrow approach to similarity of fRM, only morphosyntactic features triggering movement enter into the calculation of similarity. Under the broad approach to similarity of SIMM, a wider range of syntactic, semantic and phonological features can modulate similarity. Three main lines of empirical evidence characterize the results.

First, the results show that animacy, reversibility and lexical restriction all influence the acceptability of wh-islands. Experiment 1 revealed that the acceptability of wh-islands was modulated by similarity in animacy: higher acceptability ratings were found for sentences containing wh-elements differing in animacy, where the wh-object was inanimate and the wh-subject was animate, than for those in which both wh-elements were animate. Results from Experiment 2 revealed that the acceptability of wh-islands was modulated by thematic role reversibility, a set of semantic features that link the arguments to the verb: higher acceptability ratings were found for non-reversible sentences than for sentences in which the two arguments were similar in that they could both be the agent and the patient of the action. Results of Experiments 1, 2 and 6 showed that lexical restriction, a morphosyntactic feature with semantic correlates, also plays a role, with conditions containing two lexically restricted, semantically distinct wh-elements receiving higher ratings than those involving two bare, similar wh-words. These results suggest that the catalogue of features modulating similarity extends beyond features triggering movement, contra the narrow approach to similarity and in line with the broad approach to similarity.

Second, the data show that animacy, reversibility and lexical restriction also affect the acceptability of fully grammatical sentences. Experiments 3 and 4 revealed, with a two-alternative forced-choice task, that animacy and reversibility had a similar effect on wh-islands and on fully grammatical that-clause sentences. Experiment 6, based on an acceptability judgment task on a Likert scale, showed that lexical restriction also affects the acceptability of that-clauses, replicating previous results in English (Goodall 2015). These results show that similarity affects long-distance dependency independently of whether the sentence is acceptable (that-clauses) or not (wh-islands).
Third, observations suggest that similarity in the +Q feature has a special role to play in sentence acceptability. Experiments 5, based on a two-alternative forced-choice procedure, reported a strong preference for that-clauses, characterized by a dissimilarity in terms of the +Q feature (it is only present on the object), as compared to wh-islands characterized by an overlap of the +Q feature (it is present on both the object and the subject). Importantly, the strength of that preference (97%) is substantially higher than that of similarity in terms of animacy and reversibility reported in Experiments 3 and 4 (around 60% and 70% respectively). Experiment 6, which allowed a direct assessment of the strength of these preferences, revealed that similarity in terms of the +Q feature has a significantly stronger effect than similarity in terms of lexical restriction. These results are in line with fRM, which grants a key role to features triggering movement. They are not easily accountable under SIMM in its current formulation, which does not predict any asymmetry in the degree to which different features cause similarity-based interference.

In sum, whereas some results fit the predictions of SIMM, others fit those of fRM. In order to account for this empirical picture, I propose to distinguish between two types of similarity effects: effects due to similarity in terms of features triggering movement, responsible for setting a ban on sentence grammaticality, and effects due to similarity in terms of other linguistic features (syntactic, semantic and phonological), playing a role in modulating the memory mechanisms underlying sentence processing. I discuss these effects in the next two sections.

### 3.10.1 Similarity effects in the grammar

A major turn has been adopted in the recent development of fRM, which consists in extending the explanatory range of the theory beyond defining the boundaries of grammar, to accounting for variations in the processing difficulty of fully grammatical structures (see also section 1.2.3). In particular, fRM has been argued to account for why children fail to parse object relative clauses and adults still show difficulties (Friedmann et al. 2009, Belletti et al. 2012). Although fRM still grants a key role to features triggering movement, as was the case in the early formulation of the theory (Rizzi 2001, 2004), it extends the set of features relevant for intervention effects. In particular, lexical restriction is introduced into the catalogue of relevant features, under the assumption that it triggers movement when accompanied by a criterial
feature as +Q and +Rel. This modification allowed accounting for observations about object relative clause processing (e.g., Friedmann et al. 2009, Adani et al. 2010, Belletti et al. 2012; see Bentea et al. 2016 for a discussion) as well as wh-island processing (Experiments 1-3, Chapter 2).

Nevertheless, this new formulation of fRM is challenged by various empirical findings, revealed in previous studies as well as the current study, which I summarize here because they motivate my proposal. First, wh-islands with two lexically restricted elements (+Q, +N…+Q-N), argued to instantiate a case of inclusion (+Q, +N…+Q, Rizzi 2011), are judged more acceptable than wh-islands with a restricted extractee and a bare intervener, instantiating the same inclusion set (+Q, +N…+Q) (converging evidence in English: Atkinson et al. 2015; French: Villata et al. 2015, Villata et al. 2016). Second, although wh-islands involving feature inclusion (+Q, +N…+Q) are rated better than those involving feature identity (+Q…+Q), they remain very low, although they are considered to be above the cut-off point of adult grammar, assumed to be between feature identity and inclusion by fRM (Experiments 1, 2 and 6 of the current study, and Experiments 1-3 of Chapter 2; see also Atkinson et al. 2015, Goodall 2015, Villata et al. 2016). Third, this transition across the assumed cut-off point of adult grammar, between identity and inclusion, only shows a small difference in acceptability ratings, whereas the transition across two configurations situated within the grammatical side of the cut-off point, criterial inclusion (+Q, +N…+Q) and non-criterial inclusion (+Q, +N…+N) show big differences in acceptability ratings. This point is illustrated in Figure 6, where the dotted line represents the boundary of adult grammar according to fRM. Forth, similarity in terms of features that do not trigger movement, like animacy and thematic role reversibility, affects the acceptability of wh-islands to a similar extent as lexical restriction (d = 0.3 for lexical restriction, d = 0.24 for animacy and d = 0.1 for reversibility), although only lexical restriction is assumed to be relevant to intervention effects under fRM. Fifth, similarity in terms of such features also affects the processing of the corresponding fully grammatical structures: similarity in lexical restriction affects the acceptability of that-clauses (Experiment 6; see also Goodall 2015) and the comprehension of object relatives (Friedmann et al. 2009, Gordon et al. 2001); similarity in animacy affects the acceptability of that-clauses (Experiment 3) and the comprehension of object relatives (Corrêa 1995, Arosio et al. 2010, Bentea and Durrleman 2014, Bentea et al. 2016, Gennari and MacDonald 2008, Mak et al. 2002, 2006, Traxler et al. 2002, 2005); similarity in thematic role reversibility affects the acceptability of that-clauses
(Experiment 4) and the comprehension of object relatives (Caramazza and Zurif 1976, van der Lely and Harris 1990, Stavrakaki 2001, and Garraffa and Grillo 2008). Sixth, number and gender features, which were argued to affect the comprehension of object relatives due to fRM-type intervention effects (Belleti et al. 2012), also affect their comprehension when these features fail to trigger movement, a point that I will discuss in length in Chapter 4 (see also Villata and Franck, 2016, submitted), and even when there is no intervention, as is the case of subject relatives, which were shown to be understood better when the subject and the object mismatch in agreement features (Adani 2012; Belletti et al. 2012; Adani et al. 2014; Villata and Franck 2016, submitted). In line with this, NP-type effects, in which object relative clause comprehension is facilitated when the subject is pronominalized, which were also argued to stem from intervention effects under fRM (Friedmann et al. 2009) were also found in subject relative clause comprehension, in both children (Arnon 2010) and adults (Gordon et al. 2001). Although the effects reported in subject relatives are often smaller than those in object relatives (but not always, see Adani 2012 and Adani et al. 2014), the critical point is that they do arise, independently of intervention.

In order to account for this set of facts, I hypothesize that two different classes of similarity effects are at play: discrete effects involving features triggering movement, stemming from properties of the grammar, and gradient effects involving other linguistic features, stemming from processing constraints from the memory system. Similarity effects involving features triggering movement underlie fRM-type intervention effects, and arise in the precise configuration in which a c-commanding element intervening on a long-distance dependency has identical features, giving rise to ungrammaticality. This is the case with wh-islands, and possibly other types of weak islands, like negative islands and factive islands. On this view, intervention effects are restricted to features triggering movement, and fRM is a theory defining the boundary of grammar. The second class of similarity effects involves non-criterial syntactic features like animacy, agreement features, as well as semantic features like lexical restriction and the set of features underlying thematic role reversibility. At the empirical level, the interference effects arising from similarity in terms of these non-criterial features differ from fRM-type effects in two major respects. First, they are much reduced, in that they only mildly affect sentence acceptability or comprehension (at least in the case of adults, where a direct comparison between criterial and non-criterial features is available). Second, they arise independently of whether the sentence is grammatical (like that-clauses or object relatives) or
ungrammatical (like wh-islands), and whether the sentence involves intervention (wh-islands, that-clauses, object relatives) or not (subject relatives). It therefore seems reasonable to assume that these effects do not lie in the properties of the grammar, but rather in the workings of the memory mechanisms underlying sentence processing (which I discuss in the next section). In this view, wh-islands are considered ungrammatical, because they involve two elements that are identical with respect to features triggering movement, and the ameliorative effect that we observe for these structures associate with distinctivity in terms of lexical restriction, thematic role reversibility or animacy is due to extra-grammatical processing factors. I consider these factors and the mechanisms that underlie their effect in the next section.

3.10.2 Similarity effects in processing

Sentence processing involves building the structure, and in particular, assigning syntactic roles to the verb’s arguments. It has been argued that memory plays a key role in sentence processing, through two major processes. The first process is encoding and is responsible to store the content of the lexical units in memory. Evidence suggests that the more specified the unit is, the more costly it is to encode it. For example, nouns take longer to read when they are more richly specified (e.g., the victorious four-star general in The congressman interrogated the victorious four-star general who a lawyer for the White House advised to not comment on the prisoners) than when they are less specified (e.g., general in The congressman interrogated the general who a lawyer for the White House advised to not comment on the prisoners; Hofmeister and Vasishth 2014). This shows that encoding additional features represents an additional computational cost for the parser (length was controlled for in statistical analyses). The second process is retrieval, at play when an argument of the verb, which is situated far from it, is reactivated from memory in order to link it to the verb and build the sentence structure. Evidence shows that this process operates on the basis of cues carried by the verb, serving as pointers to the long-distance element (e.g., McElree et al. 2003, Lewis and Vasishth 2005). Both these processes have been shown to be sensitive to the presence of similar elements in the sentence. Similarity at encoding has been argued to cause feature overwriting: items sharing the same feature enter in competition such that the element losing the competition also loses the feature (e.g., Nairne 1990, Oberauer and Kliegl 2006). Similarity at retrieval has been argued to cause cue-overload when multiple items match the retrieval cues (e.g., McElree et al.
2003, Van Dyke and McElree 2006, Lewis et al. 2006). I propose that the effects of lexical restriction, reversibility and animacy reported in this Chapter also lie in these mechanisms of memory encoding and memory retrieval for structure building.

**Lexical restriction.** Wh-islands with two lexically restricted wh-elements was found to be more acceptable than those with two bare wh-elements (see also Atkinson et al. 2015, Villata et al. 2016, Experiments 1-3 in Chapter 2). Similar findings have been reported for structures with superiority violations, which are more acceptable but also faster to process when the two wh-elements are lexically restricted (e.g., *Mary wondered which book which boy read*) than when they are bare (*Mary wondered what who read*, Hofmeister et al., 2013). Although wh-islands and superiority violation differ with respect to their underlying structure (see Chapter 2; see also Mirdamadi, Villata, Shlonsky and Franck, submitted), both involve retrieving a moved wh-object. These findings thus receive a natural explanation if one assumes, following Hofmeister et al. (2013), that even though restricted wh-elements have the same syntactic form, they actually differ on several semantic dimensions that increase their distinctiveness, thus reducing similarity-based interference at encoding, retrieval or both. The same cue-based retrieval mechanism is assumed to be at play in retrieving the object of that-clauses, explaining why these structures are also improved when the two arguments are lexically restricted. It is interesting to note that these findings actually stand in contrast with the various reports that object relative clauses are harder to comprehend when the two arguments are lexically restricted (e.g., *The barber that the lawyer admired climbed the mountain*) as compared to when the object is lexically restricted and the subject is less specified, i.e., it is a pronoun (e.g., *The barber that you admired climbed the mountain*, Gordon et al. 2001), a wh-operator (as in free relatives, *Show me the one that the boy is wetting*), or an impersonal pro subject (e.g., *Show me the horse that someone is brushing*, Friedmann et al. 2009). I hypothesize that the contradictions in the influence of lexical restriction in islands/superiority violation and in object relatives stems from the respective role of memory encoding and retrieval in these two types of structures. It is reasonable to assume that whereas encoding processes are the same for the two types of structures, structure building and the related process of retrieval is harder in islands/SUV than object relatives. That is, whereas the difficulty that will manifest the most in the processing of islands/SUV lies in retrieval/structure building, the difficulty that will manifest the most in the processing of object relatives would lie in encoding. Lexically restricted elements are expected
to be more costly to encode than pronouns/wh-operators (Gibson 1998, 2000), however, evidence suggests that an element that is harder to encode is easier to retrieve further at the verb (Hofmeister and Vasishth, 2014). As a result, whereas encoding cost is the same for both types of structures, this cost is expected to show up in the processing of object relatives, penalizing comprehension when the two elements are lexically restricted, while what will mostly show up in islands/SUV is the beneficial effect, at retrieval, of having encoded a richer element. Notice that this hypothesis also accounts for the finding that children and adults also parse object relatives more easily when the object is not semantically restricted, that is, when it refers to a generic noun like animal as compared to when it refers to a set-restricted noun like zebra (on children: Goodluck 2005; on adults: Donkers, Hoeks and Stowe et al. 2013): again, this may be due to the higher cost in encoding a set-restricted element compared to a more generic one. The hypothesis that encoding cost plays a key role in object relative clause processing is also supported by a series of experiments on the role of agreement features (gender and number) in object relative clause processing that I will present in Chapters 4 and 5, which suggest that the similarity-based interference effects of these features previously reported are actually encoding effects, and not retrieval effects. Other effects reported in the literature on object relative clause processing and typically interpreted as retrieval effects may also turn out to be encoding effects (Dillon 2011; Dillon, Mishler, Slogget, and Phillips 2013), a possibility which clearly deserves more attention from future studies and that will be investigated in Chapters 4 and 5.

Thematic role reversibility. The role of thematic role reversibility has also been widely acknowledged in the comprehension of grammatical structures, in particular object relatives (see, among others, Caramazza and Zurif 1976 in typical and atypical adults, van der Lely and Harris 1990, Stavrakaki 2001 in specifically language impaired children, and Garraffa and Grillo 2008 in aphasia). All these studies show that sentences containing a subject and an object with non-reversible thematic roles are easier to comprehend than those with reversible roles. However, these studies did not tease apart the role of reversibility from the role of animacy, which systematically co-varied with it. Results presented in this Chapter provide evidence that reversibility plays a significant role in the modulation of sentence acceptability independently of animacy, since an effect of reversibility was found in sentences that contained animate noun phrases only. The present results also show that the effect of reversibility is found both in fully grammatical structures (that-clause) and in wh-islands, therefore extending previous findings
to ungrammatical structures and to grammatical structures other than object relatives. The reversibility effect also finds a natural account in terms of cue-based retrieval mechanism: in non-reversible sentences, the verb provides selective semantic cues for retrieving both the subject and the object, which is not the case in reversible sentences where both noun phrases can occupy the two argument slots. Notice that the effect here cannot be due to similarity at encoding, since the two noun phrases are identical in the reversible and non-reversible conditions: the difference lies in the verb and its semantic relation to these noun phrases. Another possibility, which is not incompatible with the former, is that reversibility also facilitates the parsing in virtue of its influence on the heuristics that the parser has been shown to rely on, since reversibility was found to also affect canonical sentences in which no retrieval mechanism is assumed to take place (e.g., Ferreira 2003). Further studies are needed to determine whether reversibility affects cue-based retrieval, heuristic formation, or both; in any case, this factor appears to play a role in the process of structure building.

**Animacy.** Finally, results presented here show that animacy also impacts both fully grammatical sentences containing an extraction out of a that-clause and wh-islands: sentences with animate subjects are more acceptable if the object is inanimate than if it is animate. These results are in line with the vast literature on object relative clauses attesting to the critical role of animacy in the comprehension of grammatical structures: the well-known advantage of subject relative clauses over object relative clauses can be neutralized when the head of the object relative is inanimate and the subject is animate (see Mak et al. 2002, 2006, Traxler et al. 2002, Gennari and MacDonald 2008 for adults, and Corrêa 1995, Arosio et al. 2010, Bentea and Durrleman 2014, Bentea et al. 2016 for children). This effect could also be explained by the role of similarity in object retrieval: the object is easier to retrieve when more distinct from the subject. However, the generalization that emerges from the literature is that similarity only affects sentences with animate subjects: no difference is found between sentences with an inanimate subject and an inanimate object and sentences with an inanimate subject and an animate object, which are both processed similarly to sentences with an animate subject and an animate object (see Mak et al. 2002, 2006; Traxler et al. 2002, 2005). This finding may be explained if one assumes that animacy actually plays a role in subject retrieval/identification in parsing: animacy is a typical property of subjects, and therefore constitutes a relevant cue to it. The report that similarity in animacy only impacts sentences with animate subjects suggests
that if the verb searches for an animate element and the element that qualifies for being the subject (because of grammatical features like its structural position and nominative case) is animate, the presence of another animate element penalizes the process, while if the element that qualifies for being the subject is inanimate, the animacy of the other element does not matter. More research is necessary to understand precisely the way cues are being used, or not, by the parser.

The proposal put forward in this chapter claims that islands are banned from the grammar, in virtue of the grammatical principle of Relativized Minimality. Some authors have argued that the degradation in acceptability ratings associated with islands is the result of a combination of processing demands in working memory that combine superadditively: when the cognitive demands exceed a certain threshold, a drop in acceptability is observed (see, amongst others, Kluender and Kutas 1993; Kluender 1998; Hofmeister, Jaeger, Sag, Arnon, Snider 2007, 2013; Hofmeister and Sag 2010). Factors with high processing demands may involve long-distance dependency processing, similarity-based interference and clause boundary processing. In support to this view, the processing of clause boundaries has been shown to involve a processing cost even in sentences without long-distance dependencies, which suggests that it is a factor of complexity as such (Kluender and Kutas 1993). Along these lines, one could argue that the reason why features triggering movement have such a strong role to play in sentence processing is that similarity in terms of these features incurs a greater processing cost. All the structures considered in the current study involve movement; an intervening element endowed with a feature triggering movement may thus be a particularly good candidate for retrieving the moved target, as it is recognized as potential filler for the gap. In this view, all similarity effects would arise from processing constraints, but effects would be modulated by the strength of the retrieval cue, such that features triggering movement, being stronger cues, would show stronger effects. Processing accounts are appealing because they reduce the explanation to factors of the same nature, i.e., processing. However, they rely on a variety of factors of different types (long-distance dependency cost, clause boundary cost, similarity cost, working memory cost, etc.), the exhaustive list of which is yet to be established (Hofmeister and Sag 2010). Moreover, some of these factors, like working memory, were argued to be irrelevant (Sprouse, Wagers and Phillips 2012; but see Hofmeister, Casasanto and Sag 2013 on the role of tasks used to test working memory) and the available evidence for the independent difficulty associated with island structures suggests that they do not always incur a processing cost (Sprouse et al. 2012,
In sum, there is no consensus concerning the general picture arising from these processing models (Sprouse, Wagers and Phillips, 2013; Wagers, 2013). It is beyond the scope of this chapter to provide a definitive explanation to the low acceptability of wh-islands. Nevertheless, this study makes a convincing case for the role of similarity in acceptability judgments across-the-board: for various types of features, and for both wh-islands and fully grammatical sentences. The hypothesis that some similarity effects lie in properties of the grammatical system whereas others lie in properties of the memory system relies on the finding of a major gap in the strength of similarity effects. I am conscious that the evidence may ultimately be compatible either with a fully grammatical or with a fully processing model of these effects, if precise assumptions could be incorporated that would account for this gap. The new perspective on wh-islands I propose here has the virtue that it captures the facts. I hope that it galvanizes new discussion on the relationship between theories of grammar and theories of processing.

3.11 Conclusions

I showed that lexical restriction, animacy and reversibility of thematic roles modulate acceptability ratings of both wh-islands and fully grammatical that-clause structures. In particular, results show that the lower the similarity between the target and the intervener, the higher the acceptability rate. However, I also showed that although these features modulate acceptability ratings, their role is much reduced compared to the role played by syntactic features triggering movement: while the former only mildly modulate acceptability ratings, the latter have a dramatic effect. I argued that neither Featural Relativized Minimality nor the Similarity-based Interference Memory Model are able to account for all the data at hand: on the one hand, fRM cannot account for the role played by lexical restriction, animacy and reversibility in modulating the acceptability of wh-islands and that-clauses, nor for the observation that extractions out of wh-islands remain poorly rated even in configurations considered as grammatical by the theory; on the other hand, SIMM cannot account for the steep drop in acceptability ratings observed for wh-islands compared to that-clauses, at least not without additional assumptions. Capitalizing on insights from both theories, I suggested that two families of similarity effects have to be distinguished: effects of similarity in syntactic
features triggering movement, lying in constraints of the grammar, and effects of similarity in other features relevant for sentence processing (like animacy, reversibility and lexical restriction), lying in the workings of the memory system. The former set of features defines the limits of grammaticality in virtue of the fRM principle which holds that the intervention of a feature triggering movement on a dependency involving the same feature blocks the grammaticality of the sentence. The second set of features operates at the level of processing, modulating the ease with which the elements of the sentence are encoded or retrieved from memory as the sentence structure is being built, whether it is grammatical or not. This approach overcomes the various challenges that fRM is confronted to. Nevertheless, it aligns with fRM in granting a key role to features triggering movement. Although the distinction between performance and competence that I propose to reintroduce here appears as the best way to account for the current state of knowledge, future work may challenge this stark cleavage, through the close collaboration of linguists and cognitive psychologists.
3.12 Appendix A

Here I report analyses on Experiment 1 and 2 on raw non-transformed scores, in order to have maximal comparability with the results of the experiments of Chapter 2. A summary of the main ratings in the 4 experimental conditions in Experiment 1 is reported in Table 3.A.1.

**Table 3.A.1.** Mean ratings on a 7 points Likert scale in the 4 experimental conditions in Experiment 1.

<table>
<thead>
<tr>
<th></th>
<th>Bare wh-elements</th>
<th>Restricted wh-elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animacy mismatch</td>
<td>2.49</td>
<td>4.09</td>
</tr>
<tr>
<td>Animacy match</td>
<td>2.71</td>
<td>3.64</td>
</tr>
</tbody>
</table>

Figure 3.A.1 shows the mean acceptability scores for the four conditions of Experiment 1. A linear mixed-effects model was fitted to the data using the lmerTest package (http://www.cran.r-project.org/web/packages/lmerTest/lmerTest.pdf) in the R software environment (R Development Core Team, 2011), with random intercepts and slopes for both subjects and items, and lexical restriction and animacy as predictors (fixed factors). The Satterthwaite approximation for degrees of freedom was used to estimate p-values. A summary of the fixed effects is reported in Table 3.A.2.

![Figure 3.A.1](image-url) **Figure 3.A.1.** Mean acceptability scores for the 4 conditions of Experiment 1. Error bars show standard errors.
Analyses revealed a main effect of lexical restriction, attesting to significantly higher ratings for sentences with restricted wh-elements (M = 3.89) than for sentences with bare wh-elements (M = 2.60) (p < 0.001). No main effect of animacy was observed (β = 0.056, t = 1.200, p = 0.236), but animacy entered into a significant interaction with lexical restriction (p = 0.003). Separate models were conducted to assess the role of animacy for bare and restricted conditions. In the restricted condition, higher ratings were found for sentences with animacy mismatch (M = 4.09) than for sentences with a match in animacy (M = 3.65) (β = 0.223, t = 3.089, p = 0.004). No effect of animacy was found in the bare condition (β = -0.111, t = -1.564, p = 0.125).

Table 3.A.2. Summary of the fixed effects for Experiment 1.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.236</td>
<td>0.146</td>
<td>22.205</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Lexical Restriction</td>
<td>0.633</td>
<td>0.083</td>
<td>7.699</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Animacy</td>
<td>0.056</td>
<td>0.046</td>
<td>1.200</td>
<td>0.236</td>
</tr>
<tr>
<td>Lexical Restriction*Animacy</td>
<td>0.167</td>
<td>0.055</td>
<td>3.060</td>
<td>.003</td>
</tr>
</tbody>
</table>

A summary of the main ratings in the 4 experimental conditions in Experiment 2 is reported in Table 3.A.3.

Table 3.A.3. Mean ratings on a 7 points Likert scale in the 4 experimental conditions in Experiment 2.

<table>
<thead>
<tr>
<th></th>
<th>Bare wh-elements</th>
<th>Restricted wh-elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-reversible</td>
<td>2.75</td>
<td>3.85</td>
</tr>
<tr>
<td>Reversible</td>
<td>2.73</td>
<td>3.50</td>
</tr>
</tbody>
</table>

Figure 3.A.2 shows the mean acceptability scores for the four conditions of Experiment 2, while a summary of the fixed effects is reported in Table 3.A.4. The same data analyses as in Experiment 1 were conducted.
Results attested to a significant main effect of lexical restriction ($p < 0.001$), with higher scores when both wh-elements are restricted ($M = 3.67$) than when they are bare ($M = 2.74$). Moreover, a main effect of reversibility was also observed ($p = 0.004$), attesting to higher scores when the sentence is non-reversible ($M = 3.3$) than when it is reversible ($M = 3.12$). Finally, a significant interaction between lexical restriction and reversibility was also observed ($p = 0.037$). Separate models were conducted to assess the role of reversibility for bare and restricted conditions. In the restricted condition, we found higher ratings for the non-reversible condition ($M = 3.85$) than for the reversible one ($M = 3.50$) ($\beta = 0.170, t = 3.047, p = 0.004$). No effect of reversibility was found in the bare condition ($M = 2.75$ vs. $M = 2.74$) ($\beta = 0.012, t = 0.346, p = 0.73$).

Table 3.A.4. Summary of the fixed effects for Experiment 2.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.210</td>
<td>0.156</td>
<td>20.468</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Lexical Restriction</td>
<td>0.465</td>
<td>0.077</td>
<td>6.034</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Reversibility</td>
<td>0.091</td>
<td>0.030</td>
<td>3.043</td>
<td>.004</td>
</tr>
<tr>
<td>Lexical Restriction* Reversibility</td>
<td>0.078</td>
<td>0.036</td>
<td>2.164</td>
<td>.038</td>
</tr>
</tbody>
</table>
3.13 Appendix B

Here I report analyses on Experiment 6 on raw non-transformed scores, in order to have maximal comparability with the results of the experiments of Chapter 2. A summary of the main ratings in the 4 experimental conditions in Experiment 2 is reported in Table 3.A.5. Figure 3.A.3 shows the mean acceptability scores for the four conditions of Experiment 6, while a summary of the fixed effects is reported in Table 3.A.6.

Table 3.A.5. Mean ratings on a 7 points Likert scale in the 4 experimental conditions in Experiment 6.

<table>
<thead>
<tr>
<th></th>
<th>That clause</th>
<th>Wh-islands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bare</td>
<td>6.47</td>
<td>2.75</td>
</tr>
<tr>
<td>Restricted</td>
<td>6.78</td>
<td>4.05</td>
</tr>
</tbody>
</table>

Results attested to a significant main effect of the type of structure ($p < .001$), attesting to higher acceptability ratings for that-clauses ($M = 6.6$) than for wh-islands ($M = 3.4$). A main effect of lexical restriction was also observed with higher scores when both elements are restricted ($M = 5.4$) than when they are bare ($M = 4.6$) ($p < .001$). Finally, a significant interaction between the type of structure and lexical restriction was attested ($p < .001$). Additional models revealed a significant effect of lexical restriction in both that-clauses ($\beta = -0.142, \ SE = 0.065, \ t = -2.175, \ p = 0.032$) and wh-islands ($\beta = -0.644, \ SE = 0.117, \ t = -5.513, \ p < .001$).

Table 3.A.6 Summary of the fixed effects for Experiment 6.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>5.037</td>
<td>0.149</td>
<td>33.621</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Type of structure</td>
<td>1.682</td>
<td>0.089</td>
<td>18.786</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Lexical Restriction</td>
<td>-0.399</td>
<td>0.065</td>
<td>-6.095</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Type of structure* Lexical Restriction</td>
<td>0.261</td>
<td>0.069</td>
<td>3.766</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
**Figure 3.A.3.** Mean acceptability scores for the 4 conditions of Experiment 6. Error bars show standard errors.
Chapter 4

Similarity-based interference in agreement comprehension and production: Evidence from French object agreement

In Chapter 3, I have distinguished among two families of features: a first family containing syntactic features triggering movement, whose effects lie in contrainstns of the grammar and lead to ungrammaticality when shared both by the target and the intervener, and a second family containing other linguistic features relevant for sentence processing, such as animacy, reversibility of thematic roles and lexical restriction, which operate at the level of processing, modulating the ease with which sentential elements are encoded or retrieved from memory as the sentence structure is being built without ultimately modifying the grammatical status of the sentence. The first aim of Chapter 4 is to systematically assess the role of features belonging to this second family in the modulation of similarity-based interference effects in the processing of grammatical long-distance dependencies. To this aim, I focus on similarity in agreement features.¹ Similarity in terms of agreement features such as gender and number should contribute in easing the encoding and/or the retrieval of elements in memory without undermining the grammatical status of the sentence. If so, we expect to observe the classical facilitatory mismatch effect due to reduced similarity-based interference when the target and the intervener have different gender or number features than when they have the same

¹ It is still a matter of debate whether the criterial approach adopted for the A-bar system also extends to the A-system (see Rizzi and Shlonsky 2007 for a discussion) and therefore if phi-features such as person, number and gender can be characterized as “non-criterial”. I will not enter in this debate in this thesis and, by capitalizing on conclusions from Chapter 3, I will assume that agreement features belong to the second family of features that are tied in the functioning of the memory system, since they do not appear to generate sentence ungrammaticality when shared both by the target and the intervener (e.g., Belletti et al. 2012).
agreement feature. However, this expectation is challenged by several recent studies reporting no mismatch effect in grammatical agreement dependencies (e.g., Wagers et al. 2009, Dillon et al. 2013, Lago et al. 2015). This report is surprising because it runs against the well-attested observation that feature mismatch reduces similarity-based interference in the processing of long-distance dependencies. Moreover, as I will discuss in length in section 4.1.2, some studies do indeed report a facilitatory mismatch effect in grammatical agreement dependencies (e.g., Franck et al. 2015). In this Chapter, I further investigate the role of agreement features in modulating similarity-based interference in grammatical long-distance dependencies such as object relative clauses with the aim to reconcile apparently contradictory findings and cast light on the mechanisms underlying the processing of agreement dependencies.

The second aim of Chapter 4 is to investigate the relationship between comprehension and production. Traditionally, language comprehension and production have been treated as largely independent systems (e.g., MacDonald, Bock, and Kelly 1993; Bock 1995; Dell, Burger, and Svec 1997; Levelt et al. 1999). However, over the last couple of decades or so, it has been suggested that these two components of the human language may be more interwoven than initially supposed and unified models assuming that comprehension and production share similar mechanisms begun to develop. In particular, it has been argued that the comprehension system would be used to monitor the production plan (e.g., Levelt 1989, Hartsuiker and Kolk 2001), while the production system would assist comprehension predicting what is coming next (e.g., Pickering and Garrod 2013, Dell and Chang 2014). Moreover, both systems must share a common grammar, which also points to the fact that they cannot be treated as completely independent systems. But if comprehension and production are closely interconnected, then one would expect interference effects to manifest similarly in comprehension and production. Agreement dependencies represent a privileged ground to address this question, since interference effects in agreement dependencies are observed both in comprehension (e.g., Wagers et al. 2009, Franck et al. 2015) and in production (e.g., Bock and Miller 1991).

I report two forced-choice experiments and two self-paced reading experiments on number and gender object-past participle agreement in French in order to systematically explore the effect of a mismatching subject in the production and comprehension of object relative clauses. Results show that the presence of a mismatching subject penalizes sentence production but improves sentence comprehension. I argue that the feature mismatch effect, despite its reverse effect in production and comprehension, is actually the result of a same phenomenon, namely
similarity-based interference, but it manifests differently because of the different processes that
the parser has to undertake in agreement comprehension and production: the process of
controller selection in agreement production, and the process of structure building in the
comprehension of complex agreement dependencies.

4.1 Introduction

Since the seminal work of Bock and Miller (1991) on agreement production, it has been shown
that an intervening element situated in the vicinity of an agreement dependency (henceforth,
*attractor*) can “attract” agreement, imposing its feature on the target. Attraction errors have for
the most part been reported in structures containing an attractor noun modifying the subject
head and mismatching it in number or gender, as in *The key to the cabinets are rusty*. However,
attraction from a preverbal object situated outside the subject phrase has also been reported in
structures involving objects in preverbal position, as in *The cabinets that the key open are on
the second floor* (Bock and Miller, 1991; Staub, 2009, 2010). Whether and how the presence of
an agreement feature mismatching the controller also affects sentence comprehension is less
clear. While early studies showed a detrimental effect of mismatch, parallel to production (e.g.,
Nicol, Forster, and Veres, 1997; Pearlmutter, Garnsey, and Bock, 1999; Pearlmutter, 2000),
more recent experiments have systematically failed to show any effect of mismatch in
grammatical sentences (e.g., Dillon, Mishler, Sloggett, and Phillips, 2013; Lago, Shalom,
Sigman, Lay, and Phillips, 2015; Tanner, Nicol, and Brehm, 2014; Tucker, Idrissi, and
Almeida, 2015; Wagers, Lau, and Phillips, 2009), while the acquisition literature reported a
facilitatory effect in children (e.g., Adani, van der Lely, Forgiarini, and Guasti, 2010; Belletti,
Friedmann, Brunato, and Rizzi, 2012), in line with a single experiment in adults (Franck,
Colonna, and Rizzi, 2015). At the empirical level, the variability in the data may be due to
differences in the methods (self-paced reading, maze task, sentence classification, eye-tracking,
event-related potential, sentence picture matching task), the language (English, French,
Spanish, Hebrew, Italian), the populations (children or adults), the features (number or gender)
or the attractor type (subject modifiers or displaced objects). At the theoretical level, as
Engelmann et al. note in a recent thorough analysis of the literature on interference in sentence
comprehension, observations showing a mismatch penalty are inconsistent with the large body
of evidence showing similarity-based interference when other types of features than agreement features are manipulated (Engelmann, Jäger, and Vasishth, submitted; see also Jäger, Engelmann, and Vasishth, 2017). However, those authors failed to discuss the acquisition data showing facilitation due to mismatch (e.g., Adani et al. 2010; Belletti et al. 2012) and treated as anomalous adult data also showing facilitation (Franck et al. 2015), although these studies do potentially support the involvement of a cue-based retrieval mechanism making use of agreement cues for retrieval.

This current state of knowledge calls for additional empirical evidence in adults, and for a theoretical framework that accounts for why the effect of feature mismatch in sentence comprehension sometimes aligns with production data in showing a mismatch penalty, and sometimes with comprehension data on non-agreement long-distance dependencies in showing a mismatch facilitation (e.g., Van Dyke and Lewis, 2003; Van Dyke and McElree, 2006; Van Dyke, 2007; Van Dyke and McElree, 2011).

Here, I present data on the effect of number and gender mismatch in both sentence comprehension and sentence production, collected on the same material and the same participants in order to have a solid empirical basis for theoretical elaboration. I first briefly summarize the literature in sentence production and the theoretical account for mismatch penalty effects. I then briefly review the literature in comprehension, and lay out a proposal about the cause of the discrepancy in the data. I suggest that sentence structure plays a key role in determining the varying effects of feature mismatch observed in sentence comprehension, due to the different contribution of two components: a forward predictive component at play in simple, predictable structures and penalized by feature mismatch, and a backward component responsible for structure building, at play in complex structures involving movement and benefitting from feature mismatch. The experimental work presented here focuses on the latter, i.e., complex structures, as I explored object relative clauses. In particular, I investigated an agreement dependency that has seldom been studied: object-past participle agreement. In French, the past participle agrees in number and gender with the preverbal object, which allowed us to explore the effect of mismatch in these two features on the same agreement dependency. Experiments 1 and 2 explore the role of number and gender mismatch in production, whereas Experiments 3 and 4 explore their role in comprehension. In the General discussion, I discuss the observed findings in light of the Feature and Controller selection model of agreement computation (Franck, in press) and models of structure building granting a key
role to feature similarity (see, among others, Lewis and Vasishth, 2005; Lewis et al., 2006; McElree, 2000; McElree et al., 2003; Oberauer and Kliegl, 2006).

4.1.1 Mismatch effects in production

Different types of factors modulate attraction: semantic, syntactic, morphological and morphophonological variables have all been found to modulate attraction rates across languages. The prominent Marking and Morphing model (henceforth, M&M; e.g., Bock, Eberhard, Cutting, Meyer and Schriefers, 2001; Eberhard, Cutting and Bock, 2005) proposes that number assignment is the result of two processes that apply successively, Marking and Morphing. Marking is responsible for translating notional number into a syntactically interpretable feature (singular vs. plural). Morphing is a morphosyntactic process that reconciles the syntactic feature selected during marking with the featural morphosyntactic specification coming from the lexicon. This model treats semantic effects, in which the notional representation of the subject’s number sometimes prevails over its morphological specification, as Marking errors (e.g., collectivity or distributivity effects, see Bock, Nicol and Cutting, 1999; Bock, Eberhard and Cutting, 2004; Eberhard, 1999; Haskell and MacDonald, 2003; Vigliocco, Butterworth and Garrett, 1996). Attraction effects have a different source: on some occasions, the plural feature of the attractor illegitimately percolates up the tree and contaminates the subject phrase during Morphing, triggering attraction. Percolation explains why a feature more deeply embedded in the tree has weaker attraction power than one less deeply embedded (e.g., Bock and Cutting, 1992; Franck, Vigliocco and Nicol, 2002; Nicol, Barss and Barker, 2016).

Although the M&M model accounts for a significant number of observations, it fails to account for findings attesting to the role of similarity in attraction. The Feature and Controller Selection model (F&CS; Franck, in press) offers an account of such effects. The model involves two distinct processes: Feature selection and Controller selection. Feature selection is responsible for selecting features from the functional lexicon and encoding them on the nouns. It is the locus of notional and morpho-phonological effects, which are argued to affect the stability of the nouns’ features: features that are backed-up by semantic and/or morpho-phonological correlates are more resistant to an encoding error (e.g., Haskell and MacDonald, 2003; Middleton and Bock, 2004; Vigliocco and Franck, 1999). It may also be the locus of
contamination from the attractor’s feature, through feature percolation, a point I will return to in the General Discussion. Controller selection is responsible for selecting the agreement controller in order to copy its features onto the target. Controller selection is argued to proceed on the basis of cues such that an element that contains typical controller features has more chances to be selected than a less controller-like element, in line with the hypothesis of Badecker and Kuminiak (2007). Similarity-based interference may therefore arise in the process of controller selection when the attractor is similar to the controller. Various sets of evidence suggest that semantic, morphological and syntactic similarity increase attraction rates, none of these effects being accounted for by M&M. For example, attractors that highly overlap with the semantic features of the head trigger more attraction than those that only weakly overlap, and attraction is stronger when the controller and the attractor have the same animacy feature, or when their semantic roles are reversible (Barker, Nicol and Garrett, 2001; Hupet, Fayol and Schelstraete, 1998; Thornton and MacDonald, 2003). Morphological similarity shows up through the observation that attraction specifically arises when the attractor has nominative case, like the controller, or when they both lack case marking (Badecker and Kuminiak, 2007; Hartsuiker, Antón-Méndez and van Zee, 2001; Hartsuiker, Schriefers, Bock and Kikstra, 2003; Lorimor, Bock, Zalkind, Sheyman and Beard, 2008; Malko and Slioussar, 2013). Finally, attractors that are syntactically similar to controllers in that they occupy a syntactic position of c-commanding the agreement target trigger more attraction than those that only precede it (Franck, Lassi, Frauenfelder and Rizzi 2006; Franck, Soare, Frauenfelder and Rizzi, 2010; Franck, Colonna and Rizzi, 2015).

To sum up, in line with Badecker and Kuminiak (2007), I suggest that attraction errors may reflect an error of controller selection due to similarity-based interference triggered by attractors bearing controller-like features. If similarity-based interference plays a key role in attraction, objects should significantly interfere in subject-verb agreement, given their semantic and syntactic similarity to subjects (Bock and Miller, 1991; Franck et al., 2006, 2010, 2015), but subjects should also interfere in object-verb agreement, for the very same reason.2 Subject

---

2 It is less clear how M&M would account for attraction effects in object relatives (either in subject or object agreement), since the attractor is never part of the controller phrase and its features can therefore not percolate. A different mechanism was indeed assumed by the pioneers of M&M to account for object attraction (Bock and Miller, 1991). This mechanism is
interference on object-verb agreement has seldom been studied (Santesteban, Pickering, and Branigan, 2013) and is the focus of Experiments 1 and 2 of the current study: if similarity plays a key role in the process of Controller Selection, then a significant amount of production errors is expected to arise in object-past participle agreement due to similarity-based interference from the intervening subject.

4.1.2 Mismatch effects in comprehension

Similarity-based interference has been widely attested in the comprehension of sentences involving long-distance dependencies, including subject-verb dependencies when the similarity does not bear on agreement features (Van Dyke and Lewis, 2003; Van Dyke and McElree, 2006; Van Dyke, 2007; Van Dyke and McElree, 2011). For example, the presence of a distractor element in the sentence with the same animacy feature as the subject has been found to penalize comprehension (Van Dyke, 2007; Van Dyke and McElree, 2011). Comprehension is also harder if the distractor element is structurally similar to the subject, as when it occupies a subject position (Van Dyke and Lewis, 2003; Van Dyke, 2007). These effects have been interpreted as the signature of a cue-based retrieval process: retrieving a distal element from memory is more difficult in the presence of a distracting element that also matches the retrieval cues at the verb because that element sometimes ends up being erroneously retrieved in place of the target. But similarity-based interference also arises at encoding. A number of studies reported a facilitatory effect of feature mismatch even when no retrieval cues were present at the verb. For example, the disadvantage of object relative clauses as compared to subject relative clauses is diminished (or even eliminated) when the subject and the object are of a different syntactic type (proper name vs. definite description) (Gordon, Hendrick, and Johnson, 2001; 2004). Since the verb does not subcategorize for the syntactic type of its arguments, these findings attest to encoding interference, showing up at retrieval site (see also Fedorenko, Gibson, and Rohde, 2006; Gordon, Hendrick, and Levine, 2002). Further evidence comes from the finding that syntactically and semantically complex targets (the victorious four-star
are read more slowly than simple targets (the general, word-length controlled for), but are retrieved faster at the verb (Hofmeister and Vasishth, 2014). Again, since the verb does not subcategorize for target complexity, the authors interpreted the facilitation at retrieval as a consequence of reduced similarity-based interference at encoding: the higher specificity of complex elements increases their distinctiveness from other elements, making them more active and easier to retrieve (see also Hofmeister, Jaeger, Sag, Arnon and Snider, 2007; 2013 for similar findings in ungrammatical sentences, and Kush, Johns and Van Dyke, 2015 for similar findings on the role of phonological similarity). Encoding interference has been claimed to result from a mechanism of feature overwriting, in which elements sharing a feature enter in a competition for this feature, and the element losing the competition also loses the features, therefore resulting in a less distinctive/active memory representation (Nairne, 1990; Oberauer and Kliegl, 2006).

One would also expect similarity-based interference to show up at the verb when a distracting element carries similar agreement features as the long-distance controller. The effect may arise at retrieval, if a cue-based mechanism making use of agreement features is launched to retrieve the controller, at encoding, if feature overwriting reduces the activation level of the controller due to the presence of a distractor carrying the same feature, or both. Rather, what early studies in sentence comprehension have shown is a penalizing effect of the presence of a number mismatching attractor, in line with production data: participants tend to slow down at the verb when the subject controller is singular and the distractor is plural, compared to cases where both nouns are singular (Acuña–Fariña, Meseguer, and Carreiras, 2014; Haussler, Bader, and Bayer, 2003; Nicol, Forster, and Veres, 1997; Pearlmutter et al., 1999; Pearlmutter, 2000). Nevertheless, a variety of studies also show that feature mismatch does facilitate processing when the sentence is ungrammatical. Pearlmutter et al. (1999) were the first to report that number mismatch between the head and a prepositional phrase (PP) modifier speeds up reading at the region immediately following the critical verb. This facilitatory effect of feature mismatch in ungrammatical sentences was replicated by Wagers, Lau, and Phillips (2009) in several self-paced reading experiments testing both simple sentences with PP modifiers (controlling for spillover effects of the nominal plural preceding the verb) and object relative clauses. When the verb incorrectly agreed in number with the mismatching object distractor (e.g., *The musicians who the reviewer praise so highly will probably win a Grammy), participants read the word immediately following the relative verb faster than when neither the
attractor nor the subject matched the number of the verb (e.g., *The musician who the reviewer praise so highly will probably win a Grammy). Facilitation due to mismatch in ungrammatical sentences has also been found in recent studies using self-paced reading (Lago et al., 2015; Tucker et al., 2015), eye-tracking (Dillon et al., 2013) and event-related brain potentials (Shen, Staub, and Sanders, 2013; Tanner, Nicol, and Brehm, 2014). Interestingly, and in contrast with early studies, the 10 experiments reported across these papers all failed to find any effect of number mismatch in grammatical sentences. Following Wagers et al., these authors adopted the explanation that the grammatical-ungrammatical asymmetry (i.e., the observation that a number mismatch effect only shows up in ungrammatical sentences) is due to the involvement of a cue-based retrieval mechanism of agreement checking when an agreement error is detected. If the parser finds an element that matches the number feature on the verb, this satisfies the agreement checking process, giving rise to an illusion of grammaticality, and allowing the parser to proceed faster than when no element matches the number feature on the verb.

In contrast to early studies showing a mismatch penalty in grammatical sentences and to this recent set of studies showing no effect in grammatical sentences, a number of experiments have actually pointed to a facilitatory effect of feature mismatch in the comprehension of object relative clauses. Most of these studies were conducted with children, using the sentence-picture matching task. Adani et al. (2010) observed that the comprehension of object relative clauses is improved in English and Italian speaking children when the object and the subject mismatch in number (e.g., Show me the elephant-SG that the lions-PL are washing vs. Show me the lion-SG that the elephant-SG is washing). Belletti et al. (2012) observed that the comprehension of object relative clauses improves when the object and the subject mismatch in gender in Hebrew speaking children, while no effect was found in Italian children. Italian differs from Hebrew in that the verb does not agree in gender with its subject. In line with these findings, in a self-paced reading experiment conducted on French adults’ comprehension of object relatives, Franck, Colonna, and Rizzi (2015) found faster reading times at the verb of the object relative clause when the object mismatched the number feature of the subject controller than when it matched it. All these observations find a natural explanation if the parser is making use of agreement cues to build the structure, in line with cue-based approaches to sentence processing (e.g., Van Dyke and Lewis, 2003; Van Dyke and McElree, 2006; Van Dyke, 2007; Van Dyke and McElree, 2011).
In their thorough review of the literature on the processing of subject-verb dependencies, Engelmann, Jäger and Vasishth (submitted) did not discuss children data. They concluded, on the basis of early evidence for a mismatch penalty in sentence comprehension, that other mechanisms than cue-based retrieval must be responsible for interference effects in agreement, without specifying which. That is, whereas semantic/structural cues would be used by the cue-based retrieval mechanism in the establishment of the long-distance dependency between the verb and its arguments, agreement cues would not. Jäger, Engelmann and Vasishth (2017) reached similar conclusions in their meta-analysis of similarity-based interference in the comprehension of subject-verb agreement dependencies. However, on the one hand, the authors analysed together findings from simple sentences with a PP subject modifier (Pearlmutter et al. 1999; Wagers et al. 2009, Experiments 4, 5, 6) and findings from complex sentences such as object relative clauses (Lago et al. 2015; Franck et al. 2015; Wagers et al. 2009, Experiments 2, 3), and, on the other hand, they collapsed results from studies in which both grammatical and ungrammatical sentences were tested (e.g., Lago et al. 2015; Wagers et al. 2009) and findings from the sole study in which only grammatical sentences were tested (Franck et al. 2015). In the next section, I argue that different mechanisms underlie the comprehension of simple sentences with a PP modifier and the comprehension of complex sentences involving movement (e.g., object relatives), and that a mix of grammatical and ungrammatical sentences in the materials may reduce the reliability of agreement cues, therefore introducing a potential confound that is absent in studies testing only grammatical agreement dependencies (see also Franck et al. 2015).

4.1.3 Accounting for the empirical variability: a proposal

Table 4.1 summarizes the findings reviewed. I suggest that the effects reviewed are actually all instances of similarity-based interference at play in three different computational mechanisms. A key distinction is made between structures involving movement, which I will refer to as ‘complex’ (like object relatives and object clefts), and structures involving no movement, which I will refer to as ‘simple’ (like sentences with PP subject modifiers). I describe these mechanisms in turn.
Table 4.1. Summary of the mismatch effects reported in production and comprehension studies as a function of Type of structure and Grammaticality. Complex structures differ from simple ones in that they involve movement. A detrimental effect of mismatch between the target and the interfering element manifests in increased agreement errors in production and increased processing difficulty in comprehension, while a facilitatory effect manifests in reduced errors or processing difficulty times.

<table>
<thead>
<tr>
<th>Type of structure</th>
<th>Grammaticality</th>
<th>Mismatch effect</th>
<th>Proposed Mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Production</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simple/Complex</td>
<td>n/a</td>
<td>Detrimental</td>
<td>Controller selection</td>
</tr>
<tr>
<td>Grammatical</td>
<td>Detrimental</td>
<td>Agreement prediction</td>
<td></td>
</tr>
<tr>
<td>Simple</td>
<td>Ungrammatical</td>
<td>Facilitatory</td>
<td>Agreement prediction</td>
</tr>
<tr>
<td><strong>Comprehension</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grammatical</td>
<td>Facilitatory</td>
<td>Structure building</td>
<td></td>
</tr>
<tr>
<td>Complex</td>
<td>Ungrammatical</td>
<td>Facilitatory</td>
<td>Structure building</td>
</tr>
</tbody>
</table>

Controller selection. The computational mechanism underlying attraction in sentence production is controller selection, a mechanism responsible for selecting the controller of agreement (Badecker and Kuminiak, 2007; Franck, in press). The controller is selected on the basis of controller-like features, which include syntactic cues (like c-command of the verb), morphological cues (like nominative case), semantic cues (like animacy) as well as probabilistic cues (like linear precedence, which may act as a good proxy to c-command). At the moment of encoding the verb, the controller’s features are copied on the verb. When only one element in the sentence is controller-like, the system correctly identifies this element as the controller and the agreement on the verb is correctly computed. However, controller selection is sensitive to the presence of elements with controller-like properties in the sentence, which ends up occasionally being selected as controller, giving rise to attraction due to similarity-based interference. Although errors in controller selection are actually independent of feature mismatch per se, i.e., they may arise both in match and mismatch conditions, they are only
observable in conditions of feature mismatch, which explains the detrimental mismatch effect reported in production. In the General discussion, I discuss the possibility that attraction errors may also arise from the other mechanism of feature selection.

Agreement prediction. The mechanism at play in the comprehension of simple sentences with canonical word order (e.g., typically sentences with a PP modifying the subject head) is the same mechanism of controller selection involved in production: the parser deploys a forward, predictive mechanism allowing it to anticipate upcoming materials, including verb number (e.g., Dell and Chang, 2014; MacDonald, 2013; Pickering and Garrod, 2013; Wagers et al., 2009). Whenever the parser’s prediction does not match the feature on the verb, a mechanism of agreement checking is triggered. In grammatical sentences, the verb’s feature usually meets the parser’s prediction. However, in the presence of a mismatching attractor, the parser occasionally predicts the wrong feature due to incorrect controller selection. Thus, in this small proportion of cases, the grammatical verb is actually unexpected, forcing the parser to check the cause of the unpredicted verb and thus giving rise to a penalizing effect of mismatch (e.g., Nicol et al., 1997; Pearlmutter et al., 1999; Pearlmutter, 2000). This effect is weak and therefore difficult to detect in experimental settings because controller selection errors are rare (production studies inform us that they are between 5% and 10%). In ungrammatical sentences, the verb most of the time carries an unpredicted feature; nevertheless, the presence on the distractor word of a feature matching the verb (while the controller does not) gives rise to an illusion of grammaticality, which manifests in terms of a facilitatory effect of feature mismatch (Wagers et al., 2009). This mismatch facilitation effect in ungrammatical sentences is stronger than the mismatch penalty in grammatical sentences due to the fact that the checking process is more often triggered when the sentence is ungrammatical, thus explaining the grammatical-ungrammatical asymmetry reported in the literature (e.g., Wagers et al., 2009). The timing of these effects supports the hypothesis that they arise in this additional, checking mechanism, since they are mostly found in the post-verbal region (Pearlmutter et al., 1999).

Structure building. The computational mechanism at play in the comprehension of more complex sentences involving movement is structure building. Studies that reported mismatch facilitation in grammatical sentences all involve object relative clauses in which the object has been moved pre-verbally. When the structure is particularly challenging, as it is the case for object relatives, the parser struggles in building the structure anticipatively. Thus, when
reaching the verb, the parser still has not stabilized on a specific syntactic structure; nevertheless, its arguments need to be integrated. The parser may thus capitalize on cues at the verb to retrieve its arguments and build the structure retroactively. Agreement markers at the verb potentially provide such cues, together with other semantic and structural cues previously found to affect sentence processing (e.g., being nominative, being in subject position, being animate; Van Dyke, 2007; Van Dyke and McElree, 2011). Under this hypothesis, feature mismatch contributes to reducing similarity-based interference in the cue-based process responsible to retrieve the verb’s arguments: if agreement features are retrieval cues, the presence of only one element that matches the verb’s agreement feature increases the chance that it will be retrieved as its subject compared to when the sentence also contains a distractor with the same agreement features as the subject. But feature mismatch may also ease structure building by reducing similarity-based interference at encoding. Interference effects at the level of encoding have been claimed to result of a mechanism of feature overwriting in memory models (Nairne, 1990; Oberauer and Kliegl, 2006). According to these models, items compete for overlapping features and the item that loses the competition has its feature overwritten (i.e., set to a value of zero), which corresponds to the loss of the feature itself. That item, having lost the feature, will become less distinct or less active in memory, and therefore harder to retrieve. In this view, a mismatch in arguments’ agreement features may contribute to maintain the arguments more active in memory, and therefore make them easier to retrieve. Hence, reduced interference at encoding may also facilitate structure building (see, amongst other, Gordon et al., 2001; 2004; Hofmeister and Vasishth, 2014). Under that hypothesis, retrieval and/or encoding interference underlies the facilitatory effect of mismatch found in children’s comprehension of object relatives (Adani et al., 2010; Belletti et al., 2012). The reason why the effect in adults seems more difficult to find (but see Franck et al., 2015) may be that children struggle more than adults to parse object relatives, and therefore rely more on disambiguating cues like agreement features on the verb.

In sum, I suggest that the variability in sentence comprehension data is due to the tension between two mechanisms: (i) a forward mechanism of agreement prediction deployed when the parser manages to anticipatively build the structure as the sentence unfolds, as is the case when the sentence is structurally simple: this mechanism is the locus of the mismatch penalty reported in simple grammatical sentences and of the mismatch facilitation (illusion of grammaticality) reported in ungrammatical sentences, and (ii) a backward, structure building mechanism
deployed by the parser to build the structure when it is too complex to develop predictions about agreement: this mechanism is the locus of the mismatch facilitation reported in complex sentences involving long-distance dependencies like object relatives. The latter mechanism is assumed mostly on the basis of evidence from children, while data on adults are scarce. It is the focus of Experiments 3 and 4 of the current study.

**4.2 Current study**

On the basis of a wide review of the literature on interference effects in long-distance dependencies, Engelmann et al. (submitted) concluded that although the predictions of the cue-based retrieval model are, modulo a few adjustments, validated by the majority of the studies, those on the role of agreement cues in subject-verb dependencies contradict the model’s predictions. Their conclusion capitalizes on the results showing a detrimental effect of number mismatch in grammatical sentences or no effect, which is left unexplained. The authors insist on the need for more studies directed at understanding why agreement cues fail to meet the predictions of the cue-based retrieval model (see also Jäger, Engelmann and Vasishth, 2017). By contrast, I have highlighted the existence of a literature in language acquisition pointing to a different finding from that reported in the adult studies, and consistent with the predictions of the cue-based approach (Adani et al., 2010; Adani et al., 2014; Belletti et al., 2012). I have suggested that when the parser struggles to build the sentence’s structure, as is the case when the sentence contains a complex long-distance dependency, it makes use of agreement cues as it uses semantic and other structural cues to determine syntactic dependencies in the sentence. I suggested that agreement features may also play a role in structure building by reducing similarity-based interference at encoding. The current study aims at gathering further evidence for a facilitatory effect of feature mismatch in adults’ processing of complex structures. I explored adults’ on-line comprehension of reversible object relative clauses with animate subjects and objects, which are known to be difficult to process, even for adults (see, among others, Caramazza and Zurif, 1976; Garraffa and Grillo, 2008; Stavrakaki, 2001; van der Lely and Harris, 1990).

Rather than exploring subject-verb agreement, I investigated past participle agreement with the object. In French, when the object moves pre-verbally, as it is the case for object relatives, the
past participle agrees with the object. The advantage of studying object-participle agreement rather than subject-verb agreement is that it requires both gender and number agreement on the participle, while the main verb only agrees in number with the subject. As shown in (3), the past participle (surprises-P,F) agrees in gender and number with the pre-verbal object (les danseuses-P,F). I was interested in investigating both number and gender agreement because even though most of the literature is on number, number is phonologically silent on the past participle, while gender is sometimes audible, which may make it more likely to produce a detectable effect. More generally, it seemed desirable to gather converging evidence from both number and gender.

(3) Les danseuses que le serveur a surprises buvaient un cocktail.

The dancers-P,F that the waiter-S,M has-S surprised-P,F drank-P a cocktail

I was confident in the relevance of studying subject interference in object agreement given the finding by Santesteban, Pickering, and Branigan (2013) on Basque of significant attraction from the subject in the production of object-verb agreement. These authors reported both object interference in subject agreement and subject interference in object agreement in canonical SVO sentences and in non-canonical OSV sentences with object topicalization. More errors were generally found in OSV sentences; however, in OSV sentences, the rate of object-verb agreement errors did not significantly differ from the number of subject-verb agreement errors, which led the authors conclude that similar mechanisms were at play in the two dependencies. I explored the role of feature mismatch in object agreement in both production and comprehension on the same participants and with the same material, in order to have maximally comparable data sets. The production experiments were conducted with two goals in mind. First, it was important to determine whether French speakers produce object agreement. Although object agreement respects fundamental properties of natural syntax (e.g., it is established in a Spec-Head relation inside an Agreement projection; Bošković, 1997; Kayne, 1989; see Belletti, 2006 for a discussion), it is optional in colloquial French where the past participle is often produced in its default singular, masculine form (see Belletti, 2006). Second, it seemed important to determine whether object agreement is prone to subject attraction similarly to the way subject agreement is prone to object attraction. Whereas the latter type of interference has been attested in various languages (e.g., Bock and Miller, 1991; Franck et al., 2006, 2010; Staub, 2009), the former has only been reported in one study in Basque.
(Santesteban et al., 2013). If agreement in sentence production involves a process of controller selection, we expect interference errors consisting in agreeing the past participle with the subject rather than the object, due to the controller-like features of the subject, making it similar to the controlling object. Indeed, subjects are similar to fronted objects in that they also occupy a position c-commanding the agreement target. Moreover, both the object controller and the intervening subject are grammatical subjects in the sentence (respectively of the matrix clause and of the embedded relative) and they occupy the same structural position (i.e., [Spec, TP]) in their respective clauses. Finally, in the materials, subjects were semantically similar to objects since sentences were semantically reversible and only animate noun phrases were used. Controller selection errors are expected to manifest in mismatch conditions, since erroneous agreement with the subject in match conditions is unobservable.

The comprehension experiments were conducted to determine whether French adults make use of agreement cues in the parsing of complex object relatives with reversible thematic roles. If the comprehension of these structures is driven by a mechanism of structure building sensitive to similarity-based interference, we expect configurations in which the object and the subject mismatch in their agreement features to be easier to process than configurations of feature match. As stated above, similarity-based interference can take place at encoding, at retrieval, or both. In order to gain insight about the locus of the expected similarity-based interference effect, I also tested the comprehension of subject relatives. Since the object follows the past participle in subject relatives, there is no object agreement and the past participle always remains in its masculine singular form (e.g., Les serveurs qui ont surpris les danseuses buvaient un cocktail, The waiters-PM who have surprised-SM the dancers-PL drank-P a cocktail). Moreover, since the object remains in its canonical post-verbal position no mechanism of object retrieval is expected. Therefore, if the expected feature mismatch effect in the comprehension of object relatives lies in the cue-based mechanism of object retrieval, no effect is expected in subject relatives since the agreement target contains no object retrieval cues and all elements are in their canonical position. The observation of a facilitatory effect of feature mismatch in subject relatives would therefore bring support to the hypothesis that similarity-based interference arises at encoding. Yet, a third possibility is that it arises both at encoding and retrieval; such a possibility would gain support from the observation that the effect is stronger in object relatives than in subject relatives, under the assumption that the two effects are additive.
Experiments 1 and 2 used a forced-choice response time paradigm with a Rapid Serial Visual Presentation (RSVP) procedure where sentences are rapidly and automatically presented word by word on a computer screen and participants are asked to select the verb form that correctly completes the sentence (Staub, 2009, 2010). Although this paradigm involves an initial component of comprehension (which is also present to some extent in the standard elicitation paradigm), this task has been shown to replicate the classical findings obtained with standard production techniques, namely mismatch effects, markedness effects, syntactic position effects and notional effects (Franck et al., 2015; Staub, 2009; 2010). It seems thus legitimate to assume that this paradigm taps into aspects of sentence production that are relevant in the computation of agreement (Staub, 2009).

Experiment 1 manipulates number mismatch, keeping gender constant (masculine), while Experiment 2 manipulates gender mismatch, keeping number constant (singular). Experiments 3 and 4 tested sentence comprehension with a self-paced reading procedure: Experiment 3 manipulates number mismatch, while Experiment 4 manipulates gender mismatch. I am not aware of any other experiment investigating interference effects in gender agreement in adult native speakers. Moreover, only two studies tested all four configurations of number mismatch within the same experiment (Kaan, 2002 with ERP measures and Wagers et al., 2009, Experiment 3), an investigation that I extended to gender agreement. In addition, and unlike these studies, in the two comprehension experiments, I restricted the investigation to grammatical sentences only, in order to avoid any possible detrimental influence from the presence of ungrammatical sentences that could reduce cues reliability. The same participants took part to the comprehension and the production experiments (Experiments 1 and 3 on number for one set of participants and Experiments 2 and 4 on gender for another set of participants), in order to have maximally comparable data. The comprehension experiment was always presented first such that it was not contaminated by the explicit attention drawn to agreement in the production task.

4.3 Experiment 1: Number agreement in production
4.3.1 Method

Participants. Sixty-five students at the University of Geneva participated in this experiment in exchange for course credit. They were all native French speakers and naïve with regard to the purpose of the experiment.

Materials and design. Thirty-two sets of four conditions each were used in a 2x2 factorial design, manipulating the number of the object (singular vs. plural) and the number match between the subject and the object (match, i.e., the subject and the object have the same number vs. mismatch, i.e., the subject and the object have a different number). Participants were asked to choose between two past participle forms (singular vs. plural) to complete the sentence. The dependent variables are therefore the proportion of correct responses and the reaction times associated with the response (see the Procedure section).

Only object relatives were tested and they were introduced by a deictic presentative (voilà) to make them verb-final. The noun phrases were always animate and masculine. Only nouns forming the plural regularly by adding the morpheme –s were used. Since in French object relatives the auxiliary agrees in number with the subject and the past participle agrees in number (and gender) with the object, I embedded the past participle in an infinitival clause in which the auxiliary was in the infinitive form, in order to avoid influences from the number of the auxiliary. Moreover, an adverb was added between the auxiliary and the past participle to increase the distance between the to-be-produced past participle and the finite verb of the relative (see Table 4.2). Three verbs were used to introduce the infinitival clause (‘to claim’, ‘to believe’, ‘to want’) in order to have some variability in the material. All sentences were reversible, so that the subject and the object were both as likely to be the agent or the patient of the verb. Informal judgements from three French-speakers who did not take part in the experiment were collected so to ensure the reversibility of our sentences. Example test items are presented in Table 4.2.

Eight lists were created so that each participant was presented with 16 experimental sentences (one experimental sentence per item) and 24 fillers. Fillers consisted of object relatives in which the two response options differed in subtle orthographical errors (e.g., presence or absence of accent, typos, homophony etc.) in order to maximise the chances that participants focus on
subtle contrasts as is the case for the experimental contrast, in which the plural marker is orthographically present but not realized phonologically.

**Table 4.2.** Example of item in the four experimental conditions of Experiment 1. The two past participles in the examples indicate the options among which participants were asked to make a choice.

<table>
<thead>
<tr>
<th>Object relative (OR)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Singular object</strong></td>
</tr>
<tr>
<td>Match (SS)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Mismatch (SP)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Plural object</strong></td>
</tr>
<tr>
<td>Match (PP)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Mismatch (PS)</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

**Procedure.** A forced-choice response time paradigm with rapid serial visual presentation procedure (RSVP) was used, following Staub (2009; 2010). Materials were presented on a computer screen using the E-prime software (Schneider, Eschman, and Zuccolotto, 2012). Sentences were presented segment by segment (content word plus a grammatical word when present) at a fast pace. To initiate the trial, the participant pressed the spacebar. Then a fixation cross was presented on the screen for 1 second followed by a blank screen (150 ms). The segments were then presented one by one at the centre of the screen for 250 ms each with a 150 ms interstimulus interval. After the final interstimulus interval, the two past participle forms (singular and plural) were presented on either side of the screen and participants were asked to

³ I put the asterisk in brackets for the PS and the PP conditions when the past participle does not agree with the object due to the optionality of the object-past participle agreement in colloquial French (Belletti 2006). This contrasts to the plural agreement with a singular object in the SS and SP conditions, which is truly ungrammatical. I will return on the optionality of object-past participle agreement in the Discussion section.
select the past participle form that would provide the grammatical continuation of the sentence by pressing one of the two allowed keys on the keyboard. For a random half of the items the correct response was on the left side of the screen, and for the other half it was on the right (as in Staub, 2010). There was no response deadline in order to avoid a possible trade-off between speed and accuracy. Instructions encouraged answering as quickly as possible by pressing the corresponding key. Once participants pressed the key, the next trial began. Participants had a one 1-minute break in the middle of the experiment. The whole session lasted about 10 minutes.

4.3.2 Results

Data analyses. Response times and accuracy proportions were analysed by way of (generalized) linear mixed-effects regression models with random intercepts for participants and items using the lme4 package in R (R Development Core Team, 2016; Bates et al. 2015). Raw response times greater than 4000 ms or less than 100 ms were removed prior to statistical analyses (affecting less than 2% of the data). No additional outlier removal process was performed. Response times were analysed on trials for which participants gave a correct answer and then log-transformed to normalize residuals. I always used the maximal random-effects structure by participant and by item and the random-effects structure always had the same specification as our fixed effects (e.g., if testing for the interaction, the random structure contained the interaction). Correlations between random effects were not estimated since they often cause the model with maximal random slopes to fail to converge. Analyses are therefore conservative with respect to the generalizability of the effects of theoretical interest to new participants and items (Barr, Levy, Cheepers, and Tily, 2013). P-values were calculated by way of the Satterthwaite’s approximation to degrees of freedom with the lmerTest package (Kuznetsova, Brockhoff, and Haubo Bojesen Christensen, 2016). For comparisons between means, I report 95% confidence intervals (CIs) calculated using the stats package in R (R Core Team, 2016). Error bars in graphs represent standard errors by subject means.
Accuracy. Figure 4.1 illustrates the distribution of accuracy proportions. Results showed a main effect of number match ($\beta = 0.793$, $z = 6.096$, $p < .001$), with higher accuracy for match ($M = 0.82; CI_{sup} = 0.85$, $CI_{inf} = 0.78$) than mismatch conditions ($M = 0.62; CI_{sup} = 0.66$, $CI_{inf} = 0.58$). Results also showed a main effect of the number of the object ($\beta = -1.135$, $z = -6.865$, $p < .001$), attesting to overall higher accuracy for singular objects ($M = 0.86; CI_{sup} = 0.89$, $CI_{inf} = 0.83$) than for plural objects ($M = 0.57; CI_{sup} = 0.62$, $CI_{inf} = 0.53$). Finally, results also showed a significant interaction ($\beta = -0.241$, $z = -2.161$, $p = 0.031$), which revealed that number match has a stronger effect for singular objects ($\beta = -0.969$, $z = -3.781$, $p < 0.001$; $M = 0.77$ in SP vs. $M = 0.95$ in SS, which corresponds to a difference in logits of 1.73) than for plural objects ($\beta = -0.697$, $z = -4.546$, $p < 0.001$; $M = 0.46$ in PS condition vs. $M = 0.68$ in PP condition, which corresponds to a difference in logits of .91).

Figure 4.1. Accuracy proportion in Experiment 1.

---

4 The direction of the interaction is opposite to that suggested by Figure 4.1, as shown by the parameter estimate, which is larger for singular ($\beta = -0.969$) than for plural ($\beta = -0.697$). This is because when proportions are transformed to logits, differences in the part of the scale near 0 or 1 are magnified (see Jaeger 2008).
Response times. Figure 4.2 illustrates the distribution of responses times. Analyses revealed a main effect of number match ($\beta = -0.123, t = -6.750, p < .001$), with longer response times for mismatch conditions ($M = 2449$ ms; $CIsup = 2635$ ms, $Clinf = 2262$ ms) than match conditions ($M = 1879$ ms; $CIsup = 1997$ ms, $Clinf = 1762$ ms). No other significant effect was found.

4.3.3 Discussion

Experiment 1 showed that number mismatch negatively affects agreement production accuracy, participants being less accurate in producing past participle agreement in the presence of an intervening subject mismatching the number of the object. These results replicate, in a new language, the finding of the study in Basque showing attraction effects from the subject on object-verb agreement computation (Santesteban et al., 2013). In the view of the F&CS model that I adopt here, this finding is interpreted as the manifestation of similarity-based interference from the intervening subject in the process of selecting the agreement controller, i.e., the object. Subject interference is particularly strong, above 20% (estimated by the difference between match and mismatch conditions), which can be explained by the fact that in our materials, the subject carries many features of the object controller: it c-commands the verb, it is itself the
controller of another dependency in the sentence (subject-verb agreement at the verb *disait/disaien*), it is animate, and it is a semantically plausible agent for the verb given the semantic reversibility of our sentences. Moreover, subject-past participle agreement, despite absent in our stimuli, is frequent in French (it occurs whenever the auxiliary is *être - to be*). All these factors contributed to lead the system select the subject as controller of the object agreement dependency in a significant proportion of the cases.

Experiment 1 also reported slower response times in the mismatch condition compared to the match condition, in line with Staub (2009). This finding does not fall directly from our hypothesis that the task taps into a process of controller selection, which predicts that the process should be sensitive to the similarity between the agreement controller and potential distractors in terms of controller-like features, but not in terms of agreement features, since these features do not participate in the process of feature selection. At first glance, the slowdown observed for mismatch conditions appears in line with the hypothesis of a feature percolation mechanism according to which interfering features can contaminate the representation of the controller’s feature (as assumed in the Marking and Morphing model, Eberhard et al., 2005). Nevertheless, such contamination is unexpected from the subject phrase to the object phrase, which is the test case studied here. Rather, I propose that the effect reflects the involvement of a mechanism of inner monitoring: when the controller and the attractor mismatch in number, the system engages in an additional checking mechanism to ensure that the correct controller has been selected prior to feature copying. This operation would not take place when the two noun phrases have the same number, since in this case the system cannot get wrong in copying the number features on the verb, which explains the faster response times in the number match conditions. Such a mechanism is in line with the monitoring procedure assumed in a number of psycholinguistic models of sentence production (e.g., Hartsuiker and Kolk 2001; Levelt 1983, 1989).

Results also attested to globally higher rates of correct agreement for singular object controllers than for plural ones, and this for both match and mismatch conditions. The finding that speakers tend to produce more correct agreement when the controller is singular shows a strong tendency to produce the default singular participial form (Corbett 2000, Greenberg 1963). Indeed, whereas a singular past participle was produced in 32% of the cases in the PP condition, the reverse error consisting in producing an erroneous plural is virtually never attested, as performance in the SS condition is almost at ceiling. The propensity to produce the singular,
masculine form of the past participle is well-attested in French, and has been taken as evidence that past participle agreement is optional in colloquial French (Belletti, 2006). Nevertheless, the sensitivity to attraction reported here suggests that the same mechanisms of agreement computation are at play as in subject-verb agreement, in line with the linguistic analysis showing that object agreement respects fundamental properties of natural syntax (e.g., Bošković, 1997; Kayne, 1989).

I also found more attraction from plural subjects than from singular subjects in accuracy analyses. This finding aligns with the classical report that number attraction in subject-verb agreement is stronger for plural attractors (e.g., Bock and Miller, 1991; Bock and Cutting, 1992; Fayol, Largy, and Lemaire, 1994; Vigliocco, Butterworth, and Semenza, 1995) and extends them to object-past participle agreement. This asymmetry has traditionally been accounted for in terms of the markedness of the noun in the Marking and Morphing model: whereas plural, marked nouns have the potential to percolate up the tree, singular, bearing no feature, cannot contaminate the subject number since no feature can percolate (e.g., Bock and Eberhard 1993; Eberhard 1997). The markedness asymmetry is interpreted differently under the F&CS model (Franck, in press). In that model, markedness plays a role on the process of feature selection, in that a plural feature is argued to be more stable and therefore more resistant to a feature selection error. As a result, an attractor that has incorrectly been selected as controller has more chances to impose its feature if that feature is stable, i.e., if it is plural, than if it is singular. A similar asymmetry is expected in Experiment 2 on gender, where attraction should be stronger with unmarked masculine subjects than with marked feminine subjects.

### 4.4 Experiment 2: Gender agreement in production

#### 4.4.1 Method

**Participants.** Eighty-five students at the University of Geneva participated in this experiment in exchange for course credit. They were all native French speakers and naïve with regard to the purpose of the experiment. The participants who took part in this experiment did not participate in Experiment 1.
**Materials and design.** I manipulated the gender of the object (masculine vs. feminine) and the gender match between the object and the subject (match vs. mismatch) in a 2x2 factorial design. Noun phrases were always animate and singular. The same experimental sentences as Experiment 1 were used, adapted for the present design. In particular, since in French object relatives, the auxiliary never agrees in gender with the subject, I simplified the structure of our relative clauses by removing both the infinitival clause and the adverb (see Table 4.3). Examples of experimental items are presented in Table 4.3. Eight lists were created so that each participant was presented with 16 experimental sentences (one experimental sentence per item) and 24 filler sentences identical to those in Experiment 1.

**Table 4.3.** Example of item in the eight experimental conditions of Experiment 2.

<table>
<thead>
<tr>
<th>Object relative (OR)</th>
<th>Masculine object</th>
<th>Feminine object</th>
</tr>
</thead>
<tbody>
<tr>
<td>Match (MM)</td>
<td>Voilà / le danseur/que/le serveur/a/surpris-*surprise</td>
<td>Voilà/la danseuse/que/la serveuse/a/surprise-(*surpris</td>
</tr>
<tr>
<td>Mismatch (MF)</td>
<td>Voilà / le danseur/que/le serveur/a/surpris-*surprise</td>
<td>Voilà/la danseuse/que/la serveuse/a/surprise-(*surpris</td>
</tr>
<tr>
<td>*Here’s/the dancer-MASC/that/the waiter-MASC/has/surprised-MASC-*surprised-FEM</td>
<td>*Here’s/ the dancer-MASC/that/the waiter-FEM/has surprised-MASC-*surprised-FEM</td>
<td></td>
</tr>
</tbody>
</table>

**Procedure.** The procedure was identical to Experiment 1.

**4.4.2 Results**

**Data analyses.** The same analyses conducted for Experiment 1 were conducted here.

**Results**

**Accuracy.** Figure 4.3 illustrates the distribution of accuracy proportions. Results attested to a main effect of gender match ($\beta = 0.916, z = 5.821, p < .001$), with higher accuracy for match ($M = 0.87; CI_\text{sup} = 0.89, CI_\text{inf} = 0.84$) than mismatch conditions ($M = 0.72; CI_\text{sup} = 0.75$,
The main effect of the gender of the object ($\beta = 1.553, z = 7.181, p < .001$) attested to higher accuracy for masculine objects ($M = 0.93; CI_{sup} = 0.95; CI_{inf} = 0.91$) than feminine objects ($M = 0.65; CI_{sup} = 0.69; CI_{inf} = 0.62$). Finally, the marginally significant interaction ($\beta = 0.249, z = 1.696, p = .089$) revealed that gender match effect tended to be stronger for masculine objects ($\beta = -2.098, z = -3.778, p < .001; M = 0.88, vs. M = 0.98, which corresponds to a difference in logits of 1.9) than for feminine objects ($\beta = -1.043, z = -4.435, p < .001; M = 0.55 vs. M = 0.75, which corresponds to a difference in logits of 0.9).

![Accuracy proportions in Experiment 2.](image)

**Figure 4.3.** Accuracy proportions in Experiment 2.
Response times. Figure 4.4 illustrates the distribution of responses times. Analyses revealed a main effect of gender match ($\beta = -0.08$, $t = -6.070$, $p < .001$), with slower response times for mismatch ($M = 2033$ ms; $CIsup = 2668$ ms, $Clinf = 2395$ms) than match conditions ($M = 1744$ ms; $CIsup = 1803$ ms, $Clinf = 1685$ ms). No other significant effect was found.

Figure 4.4. Distribution of responses times (in ms) in Experiment 2.

4.4.3 Discussion

Experiment 2 showed that gender mismatch influences both accuracy and response times in the computation of past participle agreement: participants were significantly less accurate and took more time in producing the correct form of the past participle in the presence of a subject attractor mismatching the gender of the object controller. This effect replicates results from Experiment 1 on number attraction, and is in line with previous findings showing attraction in gender in other agreement dependencies (e.g., Anton-Méndez et al., 2002; Badeker and Kuminiak, 2007; Franck, Vigliocco, and Nicol, 2002; Malko and Slioussar, 2013; Meyer and Bock, 1999; Vigliocco and Franck, 1999). Results also attested to higher accuracy rates for masculine than for feminine objects, independently of the subject’s gender: I found 25% errors consisting in producing the masculine participle in the FF condition while there were virtually
no errors consisting in producing the feminine participle in the MM condition. This replicates findings from Experiment 1 in number agreement in showing that participants tend to use the default form of the participle, singular and masculine (Belletti, 2006). Finally, I found a tendency for more attraction from feminine subjects than from masculine ones, replicating the finding of Experiment 1 on number as well as findings in the literature who reported an effect of gender markedness on attraction (e.g., Badeker and Kuminiak, 2007 in Slovak; Malko and Slioussar, 2013 in Russian). Experiment 2 basically replicates Experiment 1 on number and provides further support to the hypothesis that feature mismatch negatively affects agreement production, and particularly so when the attractor is marked (e.g., feminine), in which case the feature is assumed to be more stable, therefore pushing the system to produce a feminine past participle more than a masculine attractor, in line with the F&CS model.

In sum, Experiments 1 and 2 have shown that even though object agreement gives rise to a significant amount of default agreement in spoken French, speakers consistently, although not systematically, produce it and more importantly its computation appears to follow similar mechanisms to subject-verb agreement as shown by its similar sensitivity to attraction and markedness effects. This aligns with the analysis of object agreement as a natural syntax device (e.g., Bošković, 1997; Kayne, 1989), and invalidates the analysis in terms of ‘grammatical virus’, according to which this form of agreement would be a mere epiphenomenon of language that emerges as an object of instruction (a *prestige construction*, Emonds, 1986) but that the grammar cannot fully produce (see Sobin, 1997 for a formal characterization of grammatical viruses and Mackenzie, 2013 for an application of the concept of grammatical virus to French object-past participle agreement).

This conclusion allowed us to explore the second question, which is whether French adults make use of object agreement cues on the participle to parse object relative clauses.

### 4.5 Experiment 3: Number agreement in comprehension

#### 4.5.1 Method

*Participants.* The same participants who took part to the Experiment 1 took part in this experiment.
Material and design. The material was the same as Experiment 1, although here I added subject relatives, so that the type of relative clause (subject relative, SR vs. object relative, OR) was an additional variable in our 2x2x2 factorial design with the number of the object (singular vs. plural) and the number match between the subject and the object (match vs. mismatch). Eight lists were created so that each participant was presented with 32 experimental sentences (one experimental sentence per item) and 40 filler sentences. Filler sentences were constituted of complex sentences containing movement and/or subordination, but they did not include relative clauses. Experimental sentences were decomposed into 10 regions, each containing a content word plus a grammatical word, when present. Filler sentences were decomposed in a varying number of windows, depending on their length. An example of item is given in Table 4.4.

Procedure. The experiment was programmed with E-prime. Sentences were presented on a computer screen as part of a noncumulative self-paced paradigm (Just et al. 1982). Participants were instructed to read the sentences by pressing the space bar in order to have the segments appear. Each segment was presented in the middle of the screen and disappeared as soon as the participant pressed the space bar. Each trial began with a fixation cross (400 ms) followed by an interstimulus blank screen (150 ms). Then, the segment-by-segment presentation started. Participants were told that a yes/no comprehension question would be asked at the end of each sentence. Comprehension questions specifically targeted thematic roles and were all subject questions (e.g., Did the waiter surprise the dancer? vs. Did the dancer surprise the waiter?) in order to determine whether the correct parse was built. Half the questions required a ‘yes’ answer. Since the sentences were semantically reversible, no semantic cues were available. An inter-stimulus blank screen (150 ms) separated the last window of each sentence from the corresponding comprehension question, which appeared at the centre of the screen. Trials were separated from one another by an instruction line in which we asked the participant to press the space bar as soon as he or she was ready to continue. Instructions encouraged both rapid reading and correctness in answering the question. The items were presented in a random order. Each experimental session began with four practice trials. Three pauses of 1-minute each were administrated during the task. The whole session lasted about 15 minutes.
Table 4.4. Example of item in the eight experimental conditions of Experiment 3.

<table>
<thead>
<tr>
<th></th>
<th>Object relative (OR)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Singular object</strong></td>
<td></td>
</tr>
<tr>
<td>Match (SS)</td>
<td>Le danseur/que/le serveur/disait/avoir/souvent/énervé/buvait/un cocktail/alcoolisé.</td>
</tr>
<tr>
<td></td>
<td><em>The dancer-SG/that/the waiter-SG/claimed-SG/to have/often/annoyed-SG/drank-SG/a cocktail/with alcohol.</em></td>
</tr>
<tr>
<td>Mismatch (SP)</td>
<td>Le danseur/que/les serveurs/disaiment/avoir/souvent/énervé/buvait/un cocktail/alcoolisé.</td>
</tr>
<tr>
<td></td>
<td><em>The dancer-SG/that/the waiters-PL/claimed-PL/to have/often/annoyed-PL/drank-PL/a cocktail/with alcohol.</em></td>
</tr>
<tr>
<td><strong>Plural object</strong></td>
<td></td>
</tr>
<tr>
<td>Match (PP)</td>
<td>Les danseurs/que/les serveurs/disaiment/avoir/souvent/énervés/buvaient/un cocktail/alcoolisé.</td>
</tr>
<tr>
<td></td>
<td><em>The dancers-PL/that/the waiters-PL/claimed-PL/to have/often/annoyed-PL/drank-PL/a cocktail/with alcohol.</em></td>
</tr>
<tr>
<td>Mismatch (PS)</td>
<td>Les danseurs/que/le serveur/disait/avoir/souvent/énervés/buvaient/un cocktail/alcoolisé.</td>
</tr>
<tr>
<td></td>
<td><em>The dancers-PL/that/the waiter-SG/claimed-SG/to have/often/annoyed-PL/drank-PL/a cocktail/with alcohol.</em></td>
</tr>
<tr>
<td><strong>Subject relative (SR)</strong></td>
<td></td>
</tr>
<tr>
<td>Singular object</td>
<td></td>
</tr>
<tr>
<td>Match (SS)</td>
<td>Le serveur/qui/disait/avoir/souvent/énervé/le danseur/buvait/un cocktail/alcoolisé.</td>
</tr>
<tr>
<td></td>
<td><em>The waiter-SG/who/claimed-SG/to have/often/annoyed-SG/the dancer-SG/drank/a cocktail/with alcohol.</em></td>
</tr>
<tr>
<td></td>
<td><em>The waiters-PL/who/claimed-PL/to have/often/annoyed-PL/the dancer-SG/drank-PL/a cocktail/with alcohol.</em></td>
</tr>
<tr>
<td>Plural object</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>The waiters-PL/who/claimed-PL/to have/often/annoyed-PL/the dancers-PL/drank-PL/a cocktail/with alcohol.</em></td>
</tr>
<tr>
<td>Mismatch (SP)</td>
<td>Le serveur/qui/disait/avoir/souvent/énervé/les danseurs/buvait/un cocktail/alcoolisé.</td>
</tr>
<tr>
<td></td>
<td><em>The waiter-SG/who/claimed-SG/to have/often/annoyed-SG/the dancers-PL/drank-SG/a cocktail/with alcohol.</em></td>
</tr>
</tbody>
</table>
4.5.2 Results

Data analyses. Responses to comprehension questions were analysed by way of generalized linear mixed-effects regression models, and for reading times I used linear mixed-effects regression models using the lme4 package (Bates et al., 2015) in R (R Development Core Team, 2016). Raw reading times were analysed as follows. Reading times greater than 3000 ms or less than 100 ms were removed (affecting less than 2% of the data). No additional outlier removal process was performed. Reading times were analysed only on trials for which participants gave a correct answer to the comprehension question. Indeed, since I wanted to investigate the effectiveness of agreement cues in driving the cue-based mechanism of structure building, I restricted the investigation to items for which participants built the correct parse. Reading times were log-transformed to normalize residuals and then regressed against word length and log list position (Hofmeister, 2011; Hofmeister and Vasishth, 2014), two factors that are known to affect reading times in self-paced reading tasks (i.e., longer words are associated with longer reading times and later list position with faster reading times). The residual log reading time is therefore the dependent variable analysed. All our predictive factors were dichotomous and centred by coding one level of the factor as -1 and the other as 1. As described for Experiment 1, I always used the maximal random-effects structure by participant and by item and the random-effects structure always had the same specification as our fixed effects. No correlations between random effects were estimated. When the initial models failed to reach convergence, I progressively simplified the statistical models but in no case did the model simplification result in a model with no random slopes. P-values were calculated using Satterthwaites’s approximation to degrees of freedom with the lmerTest package (Kuznetsova et al., 2016). Error bars in graphs represent standard errors by subject means.

To assess the relative clause effect, I performed two separate analyses. First, I analysed the reading times at the past participle region (annoyed; region 6 in SRs and region 7 in ORs), the critical region of interest for ORs. Second, I analysed the matrix verb region (drank; region 8 in both SRs and ORs), which represents the point at which both the subject and the object have been integrated in both SRs and ORs.
Results

Comprehension question accuracy. The distribution of mean accuracy proportions for comprehension questions is illustrated in Figure 4.5. Analyses revealed a significant effect of relative type ($\beta = 0.634, z = 31.030, p < .001$), attesting to higher accuracy for SRs ($M = 0.85; CI_{sup} = 0.86, CI_{inf} = 0.84$) than for ORs ($M = 0.68; CI_{sup} = 0.69, CI_{inf} = 0.67$), and a significant effect of number match ($\beta = -0.166, z = -8.345, p < .001$), attesting to higher accuracy for mismatching conditions ($M = 0.79; CI_{sup} = 0.79, CI_{inf} = 0.78$) than for matching conditions ($M = 0.74; CI_{sup} = 0.75, CI_{inf} = 0.73$). All other effects were non significant.

Figure 4.5. Distribution of accuracy proportions to comprehension questions in Experiment 3.

Reading times. The distribution of reading times across the different experimental conditions for both subject and object relatives is reported in Figure 4.6. I plot raw reading times in ms for the sake of readability, but statistical analyses were performed on residual log reading times.
Figure 4.6. Distribution of reading times (in ms) in the eight experimental conditions of the different regions of Experiment 3.

Past Participle region (target verb). A significant effect of relative type was found ($\beta = -0.091$, $t = -5.207$, $p < .001$), with slower reading times for ORs ($M = 529$ ms; $CI_{sup} = 565$ ms, $CI_{inf} = 493$ ms) than for SRs ($M = 383$ ms; $CI_{sup} = 398$ ms, $CI_{inf} = 367$ ms). No other significant effect was found.

Matrix verb region. No significant effect was found.

4.5.3 Discussion

Experiment 3 replicates the well-attested relative type effect, with ORs being harder to understand and slower to process at the critical relative past participle region than SRs, in line with the vast cross-linguistic literature showing that ORs involve greater processing cost than SRs (see, among others, Frauenfelder, Segui and Mehler, 1980; Gennari and MacDonald, 2008;
Gordon et al. 2001, 2004; Just and Carpenter, 1992; King and Just, 1991; Mak, Vonk and Schriefers, 2002; 2006; Schelstraete and Degand, 1998; Traxler et al., 2002). Importantly, ORs were comprehended better when the object and the subject mismatched in number than when they matched, aligning with previous findings on children’s comprehension of object relatives in English, Italian and Hebrew (Adani et al., 2010; Belletti et al., 2012). This observation also aligns with French adults’ response times in object relative clause processing which showed that, at the target verb agreeing with the subject, number mismatch between the subject and the object speeded processing as compared to when they matched (Franck et al., 2015). The finding that number mismatch facilitates object relatives’ comprehension is in line with the vast literature showing similarity-based interference in the processing of sentences with long-distance dependencies and involving other types of cues than agreement cues (e.g., Van Dyke and Lewis, 2003; Van Dyke and McElree, 2006; Van Dyke, 2007; Van Dyke and McElree, 2011). The fact that similarity in number features between the subject and the object renders sentence processing more difficult suggests that the parser, at least in some conditions, also makes use of agreement cues when retrieving the distant object and building the structure. I argue that at least two key conditions needed to be met for the mismatch facilitation to be found. One condition is that the sentence be complex enough for the parser to struggle with the building of the structure and therefore fail to predict agreement on the target. If the parser is unable to predict verb number, verb number may actually be used in the backward process of object retrieval and structure building. Another condition is that agreement cues must be reliable: when an experiment contains a high rate of sentences with agreement errors, as is the case of the various recent studies on agreement comprehension that failed to show an effect of number mismatch in grammatical sentences, the reliability of agreement cues is reduced and feature mismatch is thus expected to have a weak impact on processing (Lago et al., 2015; Wagers et al., 2009).

Unlike Franck et al. (2015), who also tested object relatives, I found no evidence for a number mismatch effect in on-line reading time measure. One possible explanation is that whereas Franck et al. tested the effect of mismatch in the processing of the subject-verb agreement dependency, I tested its effect in the processing of object-past participle agreement, which arises after verb agreement within the relative clause (disait-SG/disaint-PL). It may thus be the case that the effectiveness of the agreement cue on the past participle was diminished by the presence of agreement cues at the verb within the relative clause, which agreed with the subject and may
therefore have provided an earlier cue for structure building. However, additional analyses conducted on the verb region for ORs (disait/disaient) revealed no significant effect of number mismatch either ($t < 1$), invalidating this hypothesis. Another possibility is that whereas in Franck et al. the agreement cue was on the verb, which is an obligatory dependency in French, it was on the past participle here, that is, on a dependency that is optional in colloquial French (Belletti, 2006). Agreement cues on the past participle, being optional, may thus play a minor role than those on the verb. A third possibility is that the object relatives I used here were syntactically more complex than those used by Franck et al., since they contained an additional embedded infinitival clause. This higher complexity may have played a role in hiding the mismatch effect in on-line measures. Experiment 4 allows us to determine whether the complexity of our structures contributed to reducing the on-line effects, as it involves simpler structures: if the complexity of the structures in Experiment 3 is responsible for the lack of on-line effect of mismatch, a significant effect should be found in Experiment 4.

Finally, I also observed a mismatch effect in the comprehension accuracy of SRs. Under the hypothesis that mismatch effects reflect similarity-based interference in the process responsible to retrieve a long-distance element, no effect was expected in SRs since in these sentences, the object is in its canonical, post verbal position. It seems therefore plausible that the mismatch effect found in SRs reflects encoding interference, arising from the competition between features shared by the subject and the object at the moment of encoding or maintaining them activated (Nairne 1990, Oberauer and Kliegl 2006, Oberauer and Lange 2008). If this hypothesis is correct, a similar mismatch effect is expected if the mismatch involves gender features: encoding two nouns with the same gender should be prone to the same feature overwriting mechanism. Also, even though the object was not distant from the verb in our SRs, the subject was distant from the main verb (buvait/buvaient); it is therefore possible that the mismatch effect found in SRs reflects the involvement of a cue-based mechanism responsible to retrieve the subject, under the assumption that the subject also needs to be retrieved when the matrix verb is reached (Van Dyke and McElree, 2011). A number mismatch between the object and the subject may have facilitated subject retrieval. If the effect observed in SRs is due to cue-based subject retrieval, no mismatch effect should be observed when the mismatch involves gender features, since gender is not a feature of the target verb. If, on the contrary, an effect of gender mismatch is still observed, this would support the hypothesis that the mismatch
effect lies in the encoding process. The two hypotheses make opposite predictions, which are assessed in Experiment 4.

4.6 Experiment 4: Gender agreement in comprehension

4.6.1 Method

**Participants.** The same participants who took part in Experiment 2 took part in Experiment 4.

**Materials and design.** I manipulated the type of relative clause (subject relative vs. object relative), the gender of the object (masculine vs. feminine) and the gender match between the object and the subject (match vs. mismatch) in a 2x2x2 factorial design. The materials were the same as Experiment 2, although here subject relatives were also tested. Since the gender agreement on the past participle is sometimes audible in French, for half of the verbs it was audible (e.g., surprisMASC (*surprised*) pronounced /syʁspʁi̯/ vs. surpris-efEM pronounced /syʁspʁiz/) while for the other half it was not (e.g., saluéMASC (*greeted*) pronounced /salje/ vs. salué-efEM, also pronounced /salje/), although it was always morphologically realized through the final –e morpheme. Examples of experimental items are presented in Table 4.5. Eight lists were created so that each participant was presented with 32 experimental sentences (one experimental sentence per item) and 40 filler sentences. The same fillers used in Experiment 3 were introduced here. Experimental sentences were decomposed into 8 regions, each containing a content word plus a grammatical word, when present. Filler sentences were decomposed in a varying number of reading windows, depending on their length.
Table 4.5. Example of item in the eight experimental conditions of Experiment 4.

<table>
<thead>
<tr>
<th></th>
<th>Object relative (OR)</th>
<th>Subject relative (SR)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Masculine object</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Match (MM)</td>
<td>Le danseur/que/le serveur/a/surpris/buvait/un cocktail/alcoolisé. The dancer-MASC/that/the waiter-MASC/has/surprised-MASC/ drank/a cocktail/with alcohol.</td>
<td>Le serveur/qui/a/surpris/le danseur/buvait/un cocktail/alcoolisé. The waiter-MASC/who/has/surprised-MASC/the dancer-MASC/drank/a cocktail/with alcohol.</td>
</tr>
<tr>
<td>Mismatch (MF)</td>
<td>Le danseur/que/la serveuse/a/surpris/buvait/un cocktail/alcoolisé. The dancer-MASC/that/the waiter-FEM/has/surprised-MASC/ drank/a cocktail/with alcohol.</td>
<td>Le serveur/qui/a/surpris/la danseuse/buvait/un cocktail/alcoolisé. The waiter-FEM/who/has/surprised-MASC/the dancer-MASC/drank/a cocktail/with alcohol.</td>
</tr>
<tr>
<td><strong>Feminine object</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Match (FF)</td>
<td>La danseuse/que/la serveuse/a/surprise/buvait/un cocktail/alcoolisé. The dancer-FEM/that/the waiter-FEM/has/surprised-FEM/ drank/a cocktail/with alcohol.</td>
<td>La serveuse/qui/a/surprise/la danseuse/buvait/un cocktail/alcoolisé. The waiter-FEM/who/has/surprised-MASC/the dancer-MASC/drank/a cocktail/with alcohol.</td>
</tr>
</tbody>
</table>

Procedure. The same self-paced reading procedure used in Experiment 3 was adopted.
4.6.2 Results

**Data analyses.** The same data analyses conducted for Experiment 3 were conducted here.

**Results**

*Comprehension question accuracy.* The distribution of mean accuracy scores in the four experimental conditions is illustrated in Figure 4.7. Generalized linear mixed effect analysis revealed a significant effect of relative type ($\beta = -0.732, z = -31.447, p < .001$), with higher accuracy scores for SRs ($M = 0.878; CI_{sup} = 0.884, CI_{inf} = 0.871$) than for ORs ($M = 0.715; CI_{sup} = 0.724, CI_{inf} = 0.706$). Results also attested to a significant three-way interaction ($\beta = 0.225, z = 2.001, p = .045$). I therefore ran separate analyses for object and subject relative clauses.

![Figure 4.7](image)

**Figure 4.7.** Accuracy proportion in the comprehension question of Experiment 4.

As for ORs, results revealed a main effect of gender match ($\beta = -0.789, z = -3.834, p < .001$) attesting to higher accuracy scores for mismatching conditions ($M = 0.76; CI_{sup} = 0.77, CI_{inf} = 0.74$) than for matching conditions ($M = 0.67; CI_{sup} = 0.68, CI_{inf} = 0.65$). A significant
interaction between gender match and the gender of the object was found ($\beta = 0.774$, $z = 2.866$, $p = .004$), attesting to a significant effect of gender match for feminine objects ($M = 0.60$ in FF vs. $M = 0.78$ in FM; $\beta = -0.649$, $z = -16.427$, $p < .001$), but not for masculine objects ($M = 0.73$ in MM vs. $M = 0.74$ in MF; $\beta = -0.055$, $z = -1.390$, $p = .164$). Remaining effects were not significant.

As for SRs, results attested to a significant main effect of the gender match ($\beta = -1.198$, $z = -2.600$, $p = .009$), which revealed higher accuracy scores for mismatching conditions ($M = 0.91$; $CI_{sup} = 0.90$, $CI_{inf} = 0.92$) than for matching conditions ($M = 0.84$; $CI_{sup} = 0.83$, $CI_{inf} = 0.85$). None of the other tests was significant.

**Reading Times.** The distribution of reading times across the different experimental conditions for both subject and object relatives is reported in Figure 4.8.

![Figure 4.8](image_url)

**Figure 4.8.** Distribution of reading times (in ms) in the eight experimental conditions for the different regions of Experiment 4.
Past Participle region (target verb). A significant effect of relative type was found ($\beta = 0.054$, $t = 3.653$, $p < .001$), with slower reading times for ORs ($M = 538$ ms; $CI_{sup} = 564$ ms, $CI_{inf} = 512$ ms) than for SRs ($M = 399$ ms; $CI_{sup} = 411$ ms, $CI_{inf} = 386$ ms). Results also attested to a significant three-way interaction ($\beta = -0.025$, $t = -2.776$, $p = .010$). In order to explore the three-way interaction, I ran separate analyses for object and subject relative clauses.

For ORs, results attested to a significant effect of the gender of the object ($\beta = -0.030$, $t = -2.174$, $p = .038$), with slower reading times for feminine objects ($M = 561$ ms; $CI_{sup} = 602$ ms, $CI_{inf} = 520$ ms) than for masculine objects ($M = 516$ ms; $CI_{sup} = 548$ ms, $CI_{inf} = 484$ ms). A significant interaction between the gender of the object and gender match was also found ($\beta = -0.036$, $t = -2.273$, $p = .031$), revealing an effect of match when the object is feminine ($\beta = -0.089$, $t = -2.092$, $p = .039$), with faster reading times when the genders mismatch (FM; $M = 516$ ms) than when they match (FF; $M = 619$ ms), but not when the object is masculine ($t = 1.501$, $p = .137$) ($M = 490$ ms in MM vs. $M = 541$ ms in MF). No other test was significant.

As for SRs, no significant effect was found.

Matrix verb region. No significant effect was attested.

4.6.3 Discussion

Results from Experiment 4 replicated the well-known advantage found for SRs compared to ORs cross-linguistically and also found in Experiment 3, and this both in comprehension accuracy and in reading times at the past participle region. More importantly, data on ORs showed that a mismatch in gender between the object and the subject significantly increased comprehension accuracy and decreased reading times at the past participle region. This is in line with results from Experiment 3 on number, except that here the effect is also found in the on-line measure of reading times. This finding converges with the report of a facilitatory effect of gender mismatch in Hebrew children’s comprehension of object relatives (Belletti et al., 2012). That study reported no facilitatory effect of gender mismatch for children speaking Italian, a language in which gender is not represented on the verb, in contrast to Hebrew. Our results further support the conclusion that if gender is represented on the agreement target, gender mismatch between the object and the subject facilitates comprehension. The results
nevertheless differ from Experiment 3 in that the mismatch effect observed both in the on-line processing of ORs and in off-line comprehension questions was only found for sentences with feminine objects. Moreover, this mismatch effect is actually not due to reduced reading times and errors in the mismatch FM condition, but rather to increased reading times and errors in the match FF condition, as compared to the other three conditions FM, MF and MM which were on a par in both accuracy and reading times. I suggest that this pattern of results arises from the combination of three additive factors of difficulty: the morphological markedness of the past participle, the morphological markedness of the noun, and the gender match between the two noun phrases. First, results from Experiment 2 suggested that producing a marked feminine past participle involves a greater cost associated with the computation of the feminine feature; it is therefore plausible that a similar cost is found in comprehension when parsing a feminine past participle. Second, encoding a marked, feminine feature on the noun would also involve a cost: it has been reported that encoding the marked, plural feature on the noun has a cost that spills over onto the next reading regions (Wagers et al., 2009). Similar findings have been reported for non-agreement features, like the syntactic/semantic richness of the noun phrase (Hofmeister et al. 2007, 2013; Hofmesiter and Vasishth, 2014). Along the same lines, the presence of a feminine subject situated before the past participle could also involve a cost that spills over the past participle. Evidence in support of that hypothesis comes from the finding that indeed, encoding a feminine noun takes more time than encoding the corresponding masculine one: this effect was significant in SRs ($t = -2.255, p = .032; M = 525$ ms for feminine NPs vs. $M = 491$ ms for masculine NPs) and marginally significant in ORs ($t = -1.870, p = .062; M = 493$ ms for feminine NPs vs. $M = 428$ ms for masculine NPs). Since in this Experiment the subject and the past participle were separated only by the monosyllabic auxiliary ‘a’, the additional cost in computing the feminine subject may have spilled over onto the past participle. Finally, gender match would involve a cost due to similarity-based interference at encoding, retrieval or both. The assumption that the three factors have the same weight accounts for the observed pattern: the FF condition combines the three factors of difficulty, while in the remaining three conditions only one factor plays a role, respectively, the markedness of the participle in FM, the spillover of the markedness of the noun situated before the past participle in MF and feature match in MM, therefore accounting for the equal improvement observed for these three conditions. If markedness played a role when gender was manipulated, why did it not also play a role when number was manipulated in Experiment 3, where the same mismatch effect for singular and
plural objects was found (in the accuracy measure)? One possible reason is that the morphological realization of the plural is simpler in that it always involves the morpheme ‘s’, which is never audible (e.g., arrivé 
SG vs. arrivé 
PL (arrived), both pronounced /aʁive/), while feminine involves the morpheme ‘e’ which, when audible, has two phonological realisations, /iz/ and /it/, respectively for participles ending in –is (e.g., surprised: /syrpʁiz/ 
MASC vs. /syrpʁiz/ 
FEM) and –it (e.g., written: /ekʁit/ 
MASC vs. /ekʁit/ 
FEM). Another possible reason is that the plural marker is never audible in spoken French, while feminine is, which therefore may have reduced its salience as a cue. In any case, further investigation is required to understand that difference between gender and number. For now, I will capitalize on the main finding that the data from Experiments 3 and 4 converge in showing that feature mismatch facilitates ORs processing, which was the focus of our study.

Finally, I replicated the finding from Experiment 3 that feature mismatch not only facilitates ORs comprehension but also SRs comprehension. This invalidates the hypothesis that the number mismatch facilitation in SRs found in Experiment 3 was due to the fact that verb number was used as a retrieval cue for the subject at the main verb, since gender was not a retrieval cue in Experiment 4. It therefore seems plausible to conclude that the encoding of the noun phrases themselves is the source of the observed similarity-based interference effects found in the SRs of these two comprehension experiments (Nairne, 1990; Oberauer and Kliegl, 2006). Further evidence in support to that hypothesis comes from the finding that encoding an object of the same gender as the subject that was previously encoded is slower than encoding an object of a different gender ($M = 496$ ms vs. $M = 520$ ms; $t = 1.895$, $p = .058$). This finding is in line with what suggested by Van Dyke and McElree (2006), who pointed that observing a mismatch effect at the region in which the similar item is first encountered (i.e., the second noun phrase region) provides strong evidence for encoding interference. However, this effect was not attested for number ($t < 2$). Further evidence is therefore required before definitive conclusions can be drawn.

If encoding interference is responsible for the mismatch effect observed in SRs, one cannot exclude that it also underlies mismatch effects in ORs, a possibility that is addressed in section 4.2 of the General Discussion.

### 4.7 General Discussion
I reported four studies on number and gender object-verb agreement in French investigating the effect of feature mismatch in the production and comprehension of complex structures involving a long-distance dependency, i.e., object relatives. Results showed a consistent pattern of mismatch penalization in production (Experiments 1 and 2), aligning with classical attraction effects found in subject-verb agreement, and a reverse, consistent pattern of mismatch facilitation in comprehension (Experiments 3 and 4), in line with well-attested similarity-based interference effects found in the processing of sentences with long-distance dependencies. In the following paragraphs, I summarize my take on the role of similarity in agreement-based interference effects in production and comprehension, and then discuss a number of open questions.

4.7.1 Attraction as errors of controller selection in sentence production

Results of Experiments 1 and 2 revealed a type of attraction that had only been reported once in Basque (Santesteban et al., 2013): subject attraction on an object-verb agreement dependency. In line with the Basque study, the results reported above show that past participle agreement in French is sensitive to attraction from the subject’s features in object relatives, exactly like subject-verb is sensitive to attraction from the objects’ features in a similar structural configuration (Bock and Miller, 1991; Franck et al., 2006, 2010; Vigliocco and Nicol, 1998). Subject attraction was expected under the hypothesis that attraction arises from the incorrect selection of the agreement controller for agreement computation. I argued that the mis-selection of the subject, rather than the object, as agreement controller of part participle agreement is an instance of similarity-based interference due to the high similarity of the subject with respect to the object in terms of controller-like features (Badecker and Kuminiak, 2007). In terms of syntactic features, subjects occupy a position c-commanding the participle and actually control two other number agreement dependencies in the sentence (auxiliary and main verb agreement), which makes them good candidates as controllers for past participle agreement. At the semantic level, since the sentences were reversible, subjects were also good candidates as patients of the target verbs, and therefore as controllers for past participle agreement. It is important to note that in that view, the fact that errors arise specifically in mismatch conditions is independent of the controller selection process itself, since the same
conditions of similarity in terms of controller-like features are met in match conditions; however, they are only observable in conditions of feature mismatch.

The assumption that attraction arises from errors in controller selection is actually in a similar vein to the interpretation of object attraction errors in subject-verb agreement in object relatives originally proposed by Bock and Miller (1991). They argued that attraction in object relatives arises from a fundamentally different mechanism from attraction from PP modifiers, which is assumed to be due to the erroneous percolation of the PP’s number feature up to the subject head and its contamination. In contrast, errors in object relatives would be due to a difficulty in syntactic role assignment and, more precisely, to the incorrect assignment of the subject role to the object (e.g., it is the controller of agreement on the matrix verb, it occupies the first sentence position usually occupied by subjects). In support of their hypothesis of two different mechanisms underlying object attraction and PP attraction, the authors reported an effect of animacy in object relatives, with more attraction when the object was animate and therefore a plausible agent for the verb, while no animacy effect was found for PP sentences (see also Barker, Nicol, and Garrett, 2001). They also found that when the object was animate, participants sometimes produced incongruent completions suggesting that they indeed treated the first noun phrase as the subject of the verb. Moreover, attraction was increased when an incongruent completion was produced. Further evidence in favor of two mechanisms came from experiments conducted with the forced-choice response time paradigm. Staub (2009) showed that while, in sentences containing a PP modifier, response times did not vary with the correctness of the response, they did so in object relatives, with slower response times for incorrect productions than for correct ones. Moreover, Staub (2010) observed that whereas the effect of a number mismatching PP was distributed across trials, whether or not an error was made, the effect of a mismatching object was restricted to a few trials, on many of which an error was made. He concluded that whereas speakers’ errors in object relatives reflect confusion, another mechanism must underlie errors from PP modifiers. This mechanism may be the feature selection component of the F&CS model, under the assumption that features within the phrase can contaminate the subject feature, possibly through percolation (e.g., Bock and Eberhard 1993; Eberhard 1997; Franck et al., 2002) or feature migration (Oberauer and Kliegl, 2006; Oberauer and Lange, 2008).

Nevertheless, several studies suggest that PP attraction, like object and subject attraction in object relatives, is also sensitive to the similarity between the PP and the subject controller at
the semantic, morphological and syntactic level, with PP attractors with controller-like features triggering more attraction than those with less controller-like features (e.g., Barker, Nicol, and Garrett, 2001; Hupet, Fayol, and Schelstraete, 1998; Malko and Slioussar, 2013; Thornton and MacDonald, 2003; see section 1.1. for an exhaustive list). The observation that similarity between the head and the PP also modulates attraction from PP modifiers suggests that these errors may also result from the erroneous selection of the PP as controller. Additional evidence for controller selection errors in PP sentences comes from a recent study on the role of the attractor’s topicality in subject-verb agreement, which shows that the more topic-like the attractor is, the higher is the rate of attraction errors (Smith, Franck and Tabor, submitted).

A related question that the concept of controller selection raises is whether it involves the erroneous selection of the controller of agreement in a structure that was correctly built, or whether it involves the building of the wrong structure in which the subject occupies the object position and object occupies the subject position. The F&CS assumes the former hypothesis, according to which controller selection is a mechanism specifically dedicated to agreement computation, separate from structure building. In line with this view, Lau, Wagers, Stroud, and Phillips (2008) manipulated verb-theme plausibility in sentences containing PP attractors to determine whether the structure that underlies erroneous agreement involves the reassignment of the subject function. They manipulated the grammaticality of the sentence and the semantic plausibility of the head and attractor nouns as subject of the subsequent verb embedded in a relative clause (e.g., The phone by the toilets was/were what Patrick dialed/flushed when he…). They reasoned that if agreement attraction is limited to formal feature checking, then encountering an ungrammatical verb (were) agreeing with the attractor (the toilets) should not make the subsequent, inadequate verb (flushed) easier to process than the adequate verb (dialed) than encountering the grammatical verb (was) agreeing with the head (the phone). In contrast, if the ungrammatical verb (were) prompts the system to reassign the subject function to the attractor (the toilets), then processing the subsequent verb for which the attractor is a plausible subject (flushed) should be easier comparatively to the verb for which it is not a plausible subject (dialed). The data showed no interaction between attractor plausibility and the embedded verb, thus supporting the hypothesis that an attraction error does not give rise to a reassignment of the subject function, and therefore that agreement involves feature checking independently of structure building. However, some findings suggest that, at least in some conditions, the incorrect structure was built. I already mentioned the finding by Bock and Miller
(1991) that incongruent completions were occasionally produced in object relatives suggesting that participants mis-assigned syntactic roles. Additional empirical support for the hypothesis that controller selection is not distinct from structure building comes from the study by Santesteban et al. (2013) who had participants produce both subject and object agreement in Basque. If controller selection can go wrong while the right structure has been built, subject agreement errors and object agreement errors should arise independently of one another (because controller selection can go wrong separately for each of these two computations). In contrast, if controller selection is tied to structure building, the two agreement dependencies should go wrong simultaneously, that is, errors should be ‘double errors’ in which both subject and object agreement are erroneous. In line with the latter prediction, Santesteban et al. reported that most (but not all) attraction errors consisted in double errors in the non-canonical OSV object topicalization structures they tested (single errors were only found for subject agreement). In contrast, virtually no double errors were produced in canonical SOV structures. The preliminary generalization arising from these studies is that ‘simple’ structures that do not involve movement (sentences with PP modification in Lau et al. 2008, canonical SOV sentences in Santesteban et al., 2013) do not seem to involve errors in structure building, whereas ‘complex’ structures with movement (object relatives in Bock and Miller, 19991 and object topicalization in Santesteban et al., 2013) seem to involve the building of an incorrect syntactic structure. The possibility that some attraction errors arise on an incorrectly built structure while others arise on the correct structure may be taken as further evidence in favour of the hypothesis that two mechanisms underlie attraction: controller selection errors, arising on the incorrect structure, and feature selection errors, arising on the correct structure. This key question remains an important topic for further research.

4.7.2 Similarity-based interference from agreement features in sentence comprehension: Retrieval or encoding effects?

Results from the two comprehension experiments showed that the comprehension of object relatives containing a long-distance dependency between the object and the verb was facilitated when the intervening subject mismatched the object’s number or gender. The data align with other off-line data from the acquisition literature showing that English, Italian and Hebrew
speaking children comprehend object relatives better when the subject’s features that are morphologically realized on the verb mismatch those of the object (Adani et al. 2010, Belletti et al. 2012). Moreover, the on-line results from Experiment 4 on gender also align with the finding of Franck et al. (2015) on French speaking adults showing that reading times at the verb agreeing in number with the controller were shorter when the controller mismatched the distractor than when they matched. More generally, the facilitatory effect of feature mismatch reported here is in line with a variety of observations showing that sentence processing is easier when the verb’s arguments are more distinct from one another than when they are similar (e.g., Van Dyke, 2007; Van Dyke and Lewis, 2003; Van Dyke and McElree, 2006; Oberauer and Kliegl, 2006; Van Dyke and McElree, 2011). However, the data reported in this Chapter contrast with other reports in the adults’ literature showing that feature mismatch fails to affect the processing of grammatical sentences (e.g., Lago et al., 2015; Wagers et al. 2009) or that it even affects it in the opposite direction, that is, by penalizing processing (e.g., Nicol et al., 1997; Pearlmutter et al., 1999; Pearlmutter, 2000). I proposed that the discrepancy among experimental results can be explained if two fundamentally different mechanisms are assumed: one at play in ‘simple’ sentences and one at play in ‘complex’ sentences involving syntactic movement. In the case of simple sentences, like those containing a head and a PP modifier, I argued, in line with predictive models of parsing (e.g., Dell and Chang, 2014; MacDonald, 2013; Pickering and Garrod, 2013), that the parser engages in a forward, predictive mechanism of controller selection, identical to the one at play in production: as the parser progresses builds the structure, it selects the agreement controller head and on that basis develops an expectation for a particular verb feature. When the grammatically correct verb is encountered, it matches the parser’s prediction (most of the time), allowing the parser to move on. However, on a few occasions (from 5% to 10% of the cases, according to production studies reported in the literature), the parser predicts an erroneous verb feature under the influence of a mismatching PP, resulting in a prediction error and thus giving rise to a small, often undetected, penalty when the grammatical verb is encountered. In contrast, when the sentence is complex, as is the case with object relatives, and possibly subject relatives as well, the parser struggles with the building of the sentence. When encountering the verb, the parser has still not stabilized on a specific syntactic structure, and it therefore explores the memory space for possible verbal arguments to integrate in the structure. It initially seemed reasonable to assume, in line with much of the literature on the role of memory in sentence processing, that the key process of
structure building would be a cue-based mechanism responsible to retrieve the long-distant element (i.e., the object in object relatives). Retrieval would be launched by the target verb carrying an agreement cue pointing at the moved object, explaining the reported facilitation when the object and the distracting subject have different features.

Although similarity-based interference is the signature of interference effects arising at retrieval, three arguments suggest that retrieval may not be the mechanism underlying the feature mismatch effects observed in Experiments 3 and 4. First, most mismatch effects in agreement dependencies were reported off-line: this is what Experiments 3 and 4 here suggest, and it converges with the literature showing for the most part a lack of on-line effect in grammatical sentences (e.g., Dillon et al., 2013; Lago et al., 2015; Tanner et al., 2014; Tucker et al., 2015; Wagers et al., 2009), but a significant effect off-line (Jäger et al. 2015). Gordon et al. (2001; 2002) also reported clear off-line effects of mismatch (in other types of features) in the comprehension of object relative clauses, while the on-line effects reported at the verb could actually result of other variables (frequency and length) that were confounded in the materials (see Van Dyke and McElree 2006 for a discussion). If feature mismatch affected retrieval, its effect would be expected on-line, at retrieval site. Second, an effect of feature mismatch was also found in the comprehension of subject relatives. It is important to note that this finding is actually not new: a closer look at the acquisition studies revealed a similar facilitatory mismatch effect in subject relatives (the effect is significant in Adani (2008), Adani et al. (2014) and Belletti et al. (2012) in Hebrew, and a trend in the same direction is found in Italian in the latter). Third, the finding in Experiment 4 that gender mismatch affected the comprehension of subject relatives, although gender was not even represented on the verb, cannot have arisen at retrieval. These observations suggest that featural similarity effects do not lie in a mechanism of memory retrieval, but rather in a mechanism of memory encoding, possibly through feature overwriting (e.g., Oberauer and Kliegl, 2006). In feature overwriting models, features are represented as units that oscillate in synchrony with the other items’ features, in line with synchronization-based neuronal models (Raffone and Wolters, 2001): if an item has the features A, B and C, these features are bound together and oscillate in synchrony. Since a feature cannot oscillate in two different phases at the same time, it cannot belong to two items at the same time. As a result, competition among items with overlapping features arises and the item losing the competition will not be endowed with the feature anymore, therefore resulting in a degraded or less active memory representation. By ensuring a greater distinctiveness of the memory
representation, encoding also affects structure building: if the quality and the activation level of the memory representation are preserved, this positively impacts structure building in that it facilitates the integration of the arguments with their verb and ensures their distinctiveness in memory, therefore possibly easing subsequent recalls of the sentence (for question answering for instance).

As reviewed in the Introduction, other studies have reported similarity-based interference effects that could not have taken place at retrieval, and were therefore interpreted as encoding effects (e.g., Fedorenko et al., 2006; Gordon et al., 2001, 2002, 2004; Hofmeister and Vasishth, 2014; Jäger et al. 2015). Moreover, one cannot exclude that effects showing up at retrieval site, and thus naturally interpreted as retrieval effects, are actually encoding effects, due to the fact that retrieving an element that has not been encoded properly due to the presence of similar distractors is more difficult (e.g., Dillon 2011; Dillon et al., 2013; Van Dyke and McElree 2006). In a recent paper, Jäger Benz, Roeser, Dillon, and Vasishth (2015) conducted one self-paced reading and two eye-tracking experiments specifically designed to disentangle encoding from retrieval interference. In the two first experiments exploring the processing of reflexive-antecedent dependencies in German, the authors found higher comprehension accuracy for sentences in which the distractor and the reflexive antecedent mismatched in gender (e.g., The thief-MASC whom the dealer-FEM obliged to steal surprisingly denounced (him)self-NEUTER and the colleagues, reported the magazine) compared to when they matched (e.g., The thief-MASC whom the dealer-MASC obliged to steal surprisingly denounced (him)self-NEUTER and the colleagues, reported the magazine). Since the German reflexive (sich) is gender-neutral, and therefore provides no cue for antecedent retrieval, this effect is likely to reflect encoding interference. In the third experiment on Swedish possessives, the authors found an on-line mismatch effect for pronominal gender-marked possessives (hans-M), but not for reflexive possessives that are not gender-marked (sin). However, the effect of mismatch was detrimental, with more regressions in the mismatch condition than in the match condition. The authors nonetheless interpreted this result as evidence for cue-based retrieval, suggesting that participants actually tended to mis-retrieve the distractor. Although the data of that second experiment point to the opposite direction to the data of the first experiment and moreover require some adjustments to the cue-based retrieval model in order to account for them, the authors concluded that similarity-based interference reflects retrieval interference rather than encoding interference. Data on subject relatives reported in this Chapter, in particular those of
Experiment 4 in which gender was not a retrieval cue for the subject, provide converging
evidence to Jäger et al.’s experiments in German in showing that feature mismatch facilitates
sentence comprehension, an effect which cannot lie in a mechanism of cue-based retrieval.
If the mismatch effects found in subject relatives in Experiments 3 and 4 as well as in the
acquisition literature are encoding effects, this potentially questions the interpretation of the
status of these effects in object relatives. A key data point to answer the question is whether the
mismatch effect observed in ORs is more pronounced than the effect found in SRs (that is,
whether relative type interacts with feature match), which would suggest that whereas similarity
affects encoding in both structures, it also affects retrieval in ORs, under the hypothesis that
encoding and retrieval effects combine additively. Data from Experiment 3 show no interaction,
suggesting that number mismatch has a similar impact on both types of relatives. Results from
Experiment 4 on gender showed a stronger effect of mismatch for ORs. The acquisition studies
discussed previously are mixed: whereas some found a similar effect of mismatch in SRs and
ORs, as shown by a lack of interaction (Adani 2008; Adani et al., 2014), others found a stronger
effect in ORs (Belletti et al., 2012 in Hebrew children), or an effect of mismatch restricted to
ORs with a non-significant tendency in SRs (Belletti et al., 2012 in Italian children). The
general picture that therefore seems to emerge is that feature mismatch has an effect at encoding
in SRs and ORs, and possibly an additional effect at retrieval in ORs, although that effect is
harder to detect. In the experiments reported here, the retrieval cue was an agreement marker
on the past participle, which may be particularly weak due to the optionality of object-participle
agreement in French. This contrasts with the obligatory nature of subject-verb agreement;
indeed, the studies that showed a stronger effect of mismatch in ORs all involved subject-verb
agreement, which may provide a stronger cue (Belletti et al. 2012, Franck et al. 2015). One may
wonder whether the reason why clear evidence for retrieval interference in ORs was not found
in the current study is that some of the participants failed to represent object agreement in their
grammar. Indeed, some participants always produced default object agreement. However, the
distribution is clearly not bimodal; rather, there is a continuum from participants producing no
object agreement at all (1 in Experiment 3, 4 in Experiment 4) to participants producing it
systematically (5 and 12 respectively). Moreover, no correlation was found between
participants’ production of participle agreement (assessed by their accuracy in match
conditions) and their sensitivity to mismatch in comprehension (assessed by the difference in
accuracy between match and mismatch conditions; r(63)=0.102, p  = 0.416 for number;
This shows that the lack of a stronger effect of mismatch in ORs compared to SRs (which would have suggested the involvement of a cue-based retrieval process in ORs) is not due to the fact that some of the participants failed to represent object agreement in their grammar: if this was the case, I would have found a positive correlation between these two variables.

In sum, findings from Experiments 3 and 4 provide clear evidence for a facilitatory effect of feature mismatch in the comprehension of both object relatives and subject relatives. Similarity effects are the signature of the role of memory in sentence processing, and the results reported here more specifically support the hypothesis that similarity affected mechanisms of memory encoding at play in the process of structure building. These findings also provide evidence, although weaker, for the additional involvement of retrieval interference in object relatives. I underlined that cue-based retrieval may be weaker when the agreement cue is optional, as it is the case for the object-past participle dependency investigated here. More generally, these findings provide support for the role of encoding interference in language processing, a mechanism that has obtained much less attention in the literature than cue-based retrieval. However, effects at encoding may actually manifest at retrieval, which shows that the two components may actually be much more intertwined than expected. A finer exploration of how these two components combine in sentence processing is provided in Chapter 5.

4.8 Conclusion

On the basis of four experiments in which I investigated the production and comprehension of sentences containing object-past participle number and gender agreement in French, I reached the following conclusions. Results from production showed that a subject mismatching the object penalizes production by reducing accuracy and increasing response times, a finding that extends previous reports of attraction to a new agreement dependency. Results from comprehension, in contrast, showed that a mismatching attractor increases comprehension accuracy and, in the case of gender, decreases reading times at the critical past participle region. I suggested that similarity-based interference underlies mismatch effects observed both in production and in comprehension: mismatch effects found in production reflect similarity-based interference during the process of controller selection in agreement computation, while
mismatch effects found in comprehension reflect similarity-based interference in the process of structure building.

Since mismatch was found to impact, in a similar way, object relatives and subject relatives for which the mismatching feature was not a retrieval cue, these findings point to the key role played by encoding interference in structure building. In Chapter 5, I pursue this investigation by providing additional empirical evidence that aims at precisely investigating the mechanisms underlying the manifestation of interference effects during sentence processing.
Chapter 5

The mechanisms underlying interference effects: Evidence from Italian and English agreement

Under a parsing model that handles long-distance dependencies through cue-based retrieval (e.g., ACT-R, Lewis and Vasishth 2005), similarity-based interference can only arise as a consequence of cue-overload at retrieval. However, as shown in Chapter 4, similarity-based interference can also arise at encoding as a consequence of feature overwriting, a result that is not accounted for by current cue-based retrieval models, which do not include a mechanism for feature overwriting. In this Chapter, I extend the investigation of Chapter 4 with the aim to systematically disentangle effects due to encoding interference from effects due to retrieval interference: while in Chapter 4 I concluded for an interference effect arising at encoding based on the comparison between subject and object relatives, here I provide a more minimal comparison between structures that only differ on the agreement manipulation.

To this aim, I report two self-paced reading experiments in subject-verb gender and number agreement in Italian and English object relative clauses. In Italian, the verb does not agree in gender with the subject, therefore providing no cue for retrieval. In English, although present tense verbs agree in number with the subject, past tense verbs do not, which allowed us to test the role of number as a retrieval cue within the same language. Results from both experiments converge in showing similarity-based interference at encoding, while there is only weak evidence for an effect at retrieval. These results stand in contrast with studies emphasizing the key role of retrieval interference in the resolution of long distance dependencies. I also show that several results that have been traditionally interpreted in terms of retrieval interference are also compatible with an interpretation in terms of encoding interference. I conclude with a discussion aiming at showing how a feature overwriting mechanism responsible for the emergence of encoding interference effects could be implemented respectively in ACT-R and in a self-organized sentence processing model (SOSP).
5.1 Introduction

One characteristic property of natural language is that it allows for long-distance dependencies: elements that are not adjacent in the input may nonetheless be related to one another. Successful language comprehension thus requires non-adjacent constituents to be accessed for semantic interpretation. Object relative clauses are well-known examples of long-distance dependency: in these configurations, the internal object of the verb does not occupy its canonical post-verbal position, but it is fronted to the beginning of the clause. Upon encountering the verb of the relative clause, a successful understanding of the sentence requires the fronted object to be accessed and integrated with its verb. Re-accessing from memory previously encountered elements requires a retrieval operation. Several studies have shown that retrieval is cue-based or content-addressable, meaning that it is driven by cues that allow the parser to directly access the intended element based on its content, rather than scanning in sequence all the elements in memory (e.g., McElree and Dosher, 1989; McElree 2000, 2006; McElree, Foraker, and Dyer, 2003; Martin and McElree, 2008; Van Dyke and McElree, 2011). The retrieval cues are triggered at the integration point (e.g., the verb) and form a subset of the target’s features: only features that are cued at the verb constitute retrieval cues (e.g., if the to-be-retrieved element is feminine, but the verb does not agree in gender with the agreement controller, feminine will not be among the retrieval cues). Cue-based retrieval has been shown to be sensitive to similarity-based interference, which arises when the cues triggered by the verb are not unique to the target element, but they also resonate with other elements in memory (e.g., Van Dyke, 2002, 2006; Van Dyke and Lewis, 2003; McElree, Foraker, Dyer 2003; Lewis and Vasishth, 2005; McElree, 2006; Van Dyke and McElree, 2006, 2011). This situation is referred to as cue-overload and it is considered one of the major causes of retrieval failure: when the retrieval cues match multiple items in memory, the probability of retrieving the target decreases, therefore leading to processing difficulties (e.g., Anderson and Neely, 1996; Nairne, 2002; Watkins and Watkins, 1975; Öztekin and McElree, 2007).

However, research in the memory domain has reported another critical source of retrieval failure: feature overwriting (e.g., Nairne, 1990; Oberauer and Kliegl, 2006). Feature overwriting is supposed to occur in a certain proportion of trials during the encoding of two items in memory (see section 1.3.3). In a nutshell, the idea is that when two items share a feature, they enter in competition for that feature and the element losing the competition loses
the feature. Therefore, feature overwriting, unlike cue-overload, arises independently from the presence of a retrieval cue at the verb, as it arises during the encoding of items in memory. For this reason, encoding interference can manifest in pre-verbal regions, and in particular when the second noun phrase is encountered, precisely because this is a mechanism that operates at the level of the featural specification of the elements in memory, regardless of the featural specification of the verb (e.g., Gordon et al. 2001, 2002; Kush et al. 2015).

A consequence of feature overwriting is that the activation level of the element that has lost the competition decreases to a certain amount. Since less active elements are harder to retrieve (Lewis and Vasishth 2005), a decrease in activation also affects retrieval probabilities. As a consequence, encoding interference can also negatively impact retrieval. This means that, while an interference effect in pre-verbal regions is only compatible with encoding interference, an interference effect observed at the retrieval region (e.g., the verb) is compatible with: (i) retrieval interference, i.e., the retrieval cue at the verb matches multiple elements in memory (cue-overload), (ii) encoding interference, i.e., the target element has lost activation due to feature overwriting, or (iii) both encoding and retrieval interference. Therefore, whenever the interference effect manifests at retrieval, no firm conclusion can be drawn with respect to the root cause of the effect.

Since the localization of the effect does not allow us to discriminate between retrieval and encoding interference, I will capitalize on a critical point of difference between the two mechanisms that are supposed to underlie retrieval and encoding interference: while a necessary condition for cue-overload to manifest is that the overlapping feature is also a retrieval cue, feature overwriting arises independently of the presence of a retrieval cue at the verb. Hence, if an interference effect is observed in the absence of a retrieval cue, then it must stem from encoding interference, even if the effect is observed at the retrieval region.

In what follows, I first briefly summarize the empirical evidence for retrieval and encoding interference and the challenge to empirically disentangle the two. I then present two empirical studies on gender and number agreement conducted respectively in Italian and English with the aim to tease apart encoding and retrieval interference. To anticipate the results, clear evidence for similarity-based interference at encoding is observed, but only weak evidence for interference at retrieval is found. Moreover, the effect was more easily detectable in off-line than in on-line measures, an observation that calls for an explanation. I then conclude with a
discussion concerning how encoding and retrieval interference can be generated in current models of sentence processing.

5.1.1 Evidence for retrieval interference

Traditionally, difficulties manifesting at the region in which retrieval is supposed to be triggered have been accounted for in terms of cue-overload: when the retrieval cues resonate with multiple items in memory, the probability of a successful retrieval is lowered, thus increasing retrieval latency at the retrieval region in on-line measures and decreasing comprehension accuracy in off-line measures (e.g., Watkins and Watkins 1975; Anderson and Neely 1996). However, neither longer reading times at the retrieval region nor lower comprehension accuracy can be taken as conclusive evidence for retrieval interference, since retrieval may also be hampered as a result of encoding interference: if feature overwriting has occurred, the quality of the target representation is reduced, therefore inducing a slow down at the retrieval region and lower comprehension accuracy.\(^1\) In other words, we cannot infer that an interference effect has originated at retrieval from a difference in reading times at the retrieval region (and even less so from an effect manifesting in off-line measures).

Because of the dual source of difficulties manifesting at the retrieval region, a number of studies that have been traditionally taken as evidence for retrieval interference are actually also compatible with encoding interference. I will review these studies first and then I will discuss studies that unequivocally provide evidence for retrieval interference.

Using an eye-tracking procedure, Van Dyke (2007) showed that in structures in which the subject and the verb were separated by a relative clause (e.g., The pilot remembered that the lady who was sitting near the smelly seat/man moaned about a refund), a constituent inside a prepositional phrase embedded in the relative clause caused longer regression path times at the region following the critical verb (moaned) and lower comprehension accuracy when it was a

---

1 It remains an empirical open question which element loses the feature as a result of feature overwriting: in Nairne’s (1990) model, when feature overwriting occurs, it is the first encoded element that will have its feature overwritten; in Oberauer and Kliegl’s (2006) model, each element can have its feature overwritten with a certain probability, independently of its position.
semantically plausible subject (*man*) for the verb than when it was not (*seat*), in virtue of its animacy. Since the on-line effect was attested right after the critical retrieval region, these results were taken as evidence for semantic interference arising at retrieval. However, the same effect is predicted if feature overwriting has occurred (*man* is more similar to *lady* than *seat* in virtue of its animacy), an observation that undermines Van Dyke’s conclusion that the locus of the interference effect was at retrieval. Similar evidence was gathered with self-paced reading measures when the similarity among the target and the intervening element was syntactic: in subject-verb long distance dependencies, an intervening element occupying a subject position, as the to-be-retrieved subject, interferes more than an intervening element occupying a prepositional object position, thus generating longer reading times at the retrieval region and lower comprehension accuracy (Van Dyke and Lewis, 2003). Again, these findings cannot be taken as evidence for retrieval interference only: under the hypothesis that subjecthood is also a feature susceptible of feature overwriting, they can also result from feature overwriting.

These findings were recently specified by a study by Van Dyke and McElree (2011) suggesting that although both syntactic and semantic cues play a role in sentence comprehension, syntactic cues may serve a gating function so that only elements with matching syntactic features can generate interference at retrieval. In sentences like *The attorney who the judge realized had declared that the motion/the witness was inappropriate compromised*, the subject (*the attorney*) must be retrieved from memory to be integrated with the matrix verb (*compromised*). The authors observed greater interference for animate distractors (*the witness*) than for inanimate ones (*the motion*) only when the distractor was in subject position, like the to-be-retrieved subject, while no semantic interference was observed when the distractor occupied an object position (e.g., *The attorney who the judge realized had rejected the witness/the motion in the case compromised*). This finding suggests that structural cues (e.g., being in a subject position) may be weighted more heavily than non-structural semantic cues, such that semantic interference only arises when the distractor is a syntactically plausible element. The authors accounted for Van Dyke (2007)’s results by suggesting that distractors in adjunct position (e.g., in a prepositional phrase) may be more difficult to rule out than distractors in argument position (e.g., in an object position), as was the case in Van Dyke and McElree (2011)’s study, under the hypothesis that the syntactic features carried by adjuncts are not salient enough for the purpose of syntactic gating, therefore allowing elements carrying these features to bypass the syntactic gating making them available to produce interference.
Additional evidence that has been taken in support of retrieval interference, but that is actually also compatible with encoding interference, comes from a recent self-paced reading study on subject-verb agreement dependencies in French object relatives in adults conducted by Franck, Colonna, and Rizzi (2015). The authors reported faster reading times at the verb of the object relative when the subject and the object had different numbers as compared to when they had the same number (e.g., Jérôme speaks to the prisoner-SG/prisoners-PL that the guard-SG takes out-SG sometimes in the yard). The authors interpreted these findings in terms of retrieval interference in a cue-based retrieval architecture making use of number agreement cues at the verb for structure building operations such as retrieval. However, the same result is expected if feature overwriting has occurred when the subject and the object have the same number feature. Finally, there are also a number of studies providing off-line comprehension results in children attesting to similarity-based interference which could stem from retrieval, encoding or both. In a sentence-picture matching task with English and Italian speaking children, Adani et al. (2010) observed improved comprehension accuracy for object relative clauses when the object and the subject mismatch in number (e.g., Show me the elephant-SG that the lions-PL are washing vs. Show me the lion-SG that the elephant-SG is washing), a result that is again compatible with both retrieval and encoding interference (see also Adani et al. 2014 for similar evidence in specific language impairment children, and Bentea and Durrlemann 2014 and Bentea, Durrleman, and Rizzi 2016 for similar evidence when animacy is manipulated).

I now discuss the only two studies that, as far as I know, provide substantial evidence for retrieval interference. The first evidence comes from a study on adults. Using a memory load paradigm combined with a self-paced reading task, Van Dyke and McElree (2006) manipulated the retrieval cues at the verb such that they either do or do not uniquely identify the target in virtue of the verb’s semantic constraints (i.e., the distant object the boat) (Memory load: table, sink, truck – Sentence: It was the boat that the guy who lived by the sea sailed/fixed in two sunny days). When the retrieval cues were not unique to the target, they also matched the elements in the memory load, which were all fixable, but none of them was sailable. The authors observed longer reading times at the critical verb and lower comprehension accuracy in the cue-overload condition (fixed) than in the non-cue-overload condition (sailed), consistent with a cue-based retrieval processing sensitive to semantic similarity: when more than one element in memory matches the retrieval cues at the verb (sailable vs. fixable), interference effects arise increasing retrieval difficulties. Since the memory load was kept constant across conditions, the
observed difference can only be attributed to retrieval interference. However, despite these findings provide clear evidence for retrieval interference, they do not count as evidence against encoding interference in sentence comprehension in general, as also noted by Jäger et al. (2015). The second evidence comes from a study on children. In a sentence-picture matching task, Belletti et al. (2012) reported an improvement in comprehension accuracy for object relatives as compared to subject relative when gender was manipulated in Hebrew speaking-children, while no effect was observed for Italian children. Crucially, while in Hebrew the verb agrees in gender with the subject, therefore providing a subject retrieval cue, in Italian the verb does not agree in gender with the subject, which therefore explains the absence of the effect in Italian. Since the facilitatory effect of gender mismatch is exclusively attested when gender is a retrieval cue at the verb, these findings provide support for retrieval interference, but, again, no evidence against encoding interference.

To summarize, with the exception of studies by Van Dyke and McElree (2006) and Belletti et al. (2012), results from studies that have traditionally been interpreted as supporting retrieval interference are actually also compatible with encoding interference: if feature overwriting occurred, the quality and the activation of the memory representation are reduced and retrieval is thus hampered, thus explaining the slow down observed at the retrieval region in on-line measures and the lower comprehension accuracy observed in off-line measures.

5.1.2 Evidence for encoding interference

Evidence for feature overwriting has initially been provided by studies using the complex span paradigm in list memorization: participants are presented with a list of four words that they will be asked to recall, and each word is separated from the following by four distractors. Oberauer (2009) observed that recall was impaired when the distractors have overlapping phonemes with the target words (see Lange and Oberauer 2005 and Oberauer and Lange 2008 for similar results). The author interpreted this finding in terms of a feature overwriting mechanism: distractors with features overlapping the target’s features can overwrite the target’s feature. The authors showed that feature overwriting can be the result of pairwise similarity (i.e., the target item is highly similar to the distractor), but it can also arise in a distributed way, such that the
target has overlapping feature with several distractors (e.g., a four-letter target word that shares one letter with each of the four distractors).

The mechanism of feature overwriting has then been extended to account for interference effects observed in sentential contexts in which no retrieval cue is present at the integration region. In a series of self-paced reading experiments on relative clauses, Gordon et al. (2001, 2004) reported that the well-attested disadvantage of object relatives as compared to subject relatives was reduced or even eliminated when the subject and the object were of different syntactic kinds (e.g., a pronoun and a definite description) as compared to when they were of the same syntactic kind. Faster reading times at the verb and higher comprehension accuracy were observed in mismatch conditions (i.e., definite description vs. pronoun, The barber that you admired climbed the mountain) as compared to match conditions (i.e., two definite descriptions, The barber that the lawyer admired climbed the mountain). Since the distinction between definite description and pronoun is not a retrieval cue, the facilitation effect of mismatch observed for object relatives cannot lie in the cue-based retrieval process directed at satisfying constraints of the verb (similar results were obtained by Gordon et al. (2002) and Fedorenko et al. (2006) in a memory load paradigm). Hence, even though the interference effect was detected at the critical retrieval region (i.e., the verb), similarly to the studies reported in the previous section, these effects cannot be interpreted as being the manifestation of retrieval interference, but they are exclusively interpretable in terms of encoding interference.

Additional findings pointing at the critical role of encoding interference have also been recently provided by Hofmeister and Vasisht (2014) in a self-paced reading study. In sentences in which the to-be-retrieved object (the general) was modified by an object relative clause (e.g., The congressman interrogated the general who a lawyer for the White House advised to not comment on the prisoners), the authors observed faster reading times at the verb (advised) when the target was semantically and syntactically complex (the victorious four-star general) as compared to when it was simple (the general, word length controlled for in the statistical analyses). Since the complexity of the target is not a retrieval cue, the authors interpreted this finding as supporting encoding interference. Once again, even though the interference effect manifested at the critical verb, its root cause appears to originate at encoding. Moreover, faster reading times at retrieval were associated with longer reading times at encoding: complex targets took longer to be encoded than simple ones, but were then retrieved faster than simple ones (see also Hofmeister et al. 2007, 2013 for similar findings on ungrammatical sentences.
containing a superiority violation). The authors suggested that even though encoding additional features represents an extra computational cost for the parser, therefore resulting in longer reading times at encoding, it ensures a greater distinctiveness in memory of the target, therefore lowering the risk of feature overwriting and speeding up retrieval. Interestingly, in an additional self-paced reading experiment, the authors showed that manipulating the text color of the target element (congruent vs. incongruent with the surrounding text) had an effect at encoding, with congruent words being read faster than incongruent ones, but no effect was attested at the retrieval region. The absence of an effect at retrieval comes as no surprise under the hypothesis that the resolution of long-distance dependencies is conducted through a grammar-driven retrieval mechanism which is sensitive only to those features that are linguistically relevant for structure building operations. However, text color has an effect at encoding, with congruent matching conditions (i.e., conditions in which the target element has the same color as the surrounding text) being read faster than incongruent mismatch conditions (i.e., conditions in which the target element has a different color as the surrounding text). This result may appear surprising if one assume that feature overwriting can also occur on superficial non-linguistic dimensions such as color, since in this case we would expect match conditions to come with a disadvantage as compared to mismatch conditions, an implication that the authors do not address.

In line with these results showing that superficial features do not play any role in structure building operations such as retrieval, but can modulate encoding, Kush, Johns, and Van Dyke (2015) report effects of phonological similarity at the encoding region only. In two studies using a memory-load paradigm in a self-paced reading task, the authors manipulated the word in the memory load such that they either rhyme or not with the to-be-retrieved element (the boat) in an object cleft clause (e.g., Rhyme Memory Load: coat, vote, note; No Rhyme Memory Load: table, sink, truck; Sentence: It was the boat that the guy who drank some hot coffee sailed on two sunny days). Results attested to longer reading time at the second noun phrase region (that the guy) in the rhyme condition as compared to the no-rhyme condition, thus attesting to a detrimental effect of phonological overlap at encoding, while no effect was observed at the critical verb region, consistent with a grammar-based parser sensitive only to those features that are useful to build the structure, and contra Acheson and MacDonald (2011), who argued that phonological interference effects derive from retrieval interference, even though their results were compatible with encoding interference as well.
The interference effect found at the second noun phrase region by Kush et al. (2015) is in line with studies by Gordon et al. (2001, 2002) who also systematically found interference effects in that region. Gordon et al. observed slower reading times at the second noun phrase when it was of the same syntactic kind as the target element or the elements in the memory load (e.g., definite description) than when it was of a different syntactic kind (e.g., pronoun). However, as noted by Van Dyke and McElree (2006), the pre-verbal effects found in Gordon et al. ’s studies are not unequivocally interpretable in terms of encoding interference: in Gordon et al. (2001), pronouns are both shorter and more frequent than definite descriptions, and in Gordon et al. (2002) the interference effect was already attested in the region containing the first noun phrase, an effect that may thus be due to a rehearsal strategy adopted by participants in the attempt to distinguish items in the memory list from sentential elements. In addition, and perhaps more critically, even the effect observed at the verb cannot uniquely be attributed to encoding interference, since it could be a spillover effect from the second noun phrase region that was adjacent to the verb. These critiques, however, do not hold for the effect observed by Kush et al. (2015). Moreover, even if the interpretation of the on-line effects in Gordon et al. ’s studies appear to be questionable for the reasons above, their results from comprehension accuracy still provide support for encoding interference: when items in memory lose distinctiveness due to feature overwriting, sentence comprehension is hampered, thus lowering accuracy in comprehension questions. Since the syntactic type of the target is not a retrieval cue, this off-line result is plausibly the result of encoding interference.

Finally, mixed evidence comes from a recent study by Jäger, Benz, Roeser, Dillon, and Vasishth (2015) specifically designed to disentangle encoding and retrieval interference in reflexive dependencies. In one self-paced reading and two eye-tracking experiments, the authors tested the effect of mismatching gender features in contexts in which no gender retrieval cue was provided (i.e., German antecedent dependencies, where the reflexive, sich, is gender-neutral) and in contexts in which a gender retrieval cue was either provided (i.e., Swedish possessives, which are gender marked, hans-M) or not (i.e., Swedish reflexive possessive, which are not gender-marked, sin-Ø). The predictions are straightforward: encoding interference predicts a facilitatory effect of mismatch across the board, whereas retrieval interference predicts a facilitatory effect of mismatch only in the Swedish possessive condition, which is the only condition in which gender is a retrieval cue. Results from the three experiments can be summarized as follows: first, in the two German experiments (self-paced reading and eye-
tracking), although no online effects were detected, higher comprehension accuracy rates were found in the gender mismatch condition as compared to the match condition. Since the German reflexive (sich) is gender-neutral, therefore providing no cue for retrieval, this effect is only compatible with encoding interference. Second, for Swedish possessives, both an on-line and an off-line mismatch effect was observed for pronominal gender-marked possessives, but not for non-gender marked reflexive possessive. However, the on-line mismatch effect showed up in the opposite direction to similarity-based interference effects: more regressions were observed in the mismatch condition than in the match condition. The authors interpreted this effect as a retrieval interference effect, reflecting the mis-retrieval of the interfering element, and put aside the German results where evidence for encoding interference was attested off-line. To account for the direction of the on-line result, which is at odds with the usual reports of a mismatch facilitatory effect, the authors claimed that mis-retrieval occurs in the presence of highly active distractors: in their material, the distractor was in a particularly salient syntactic position, i.e., the subject position, and it was also linearly closer to the retrieval site, two ingredients that, according to the authors, render the intervener particularly active in memory, therefore leading to mis-retrieval. However, first, this assumption requires adjustments in the ACT-R model (Lewis and Vasishth, 2005) that, as such, does not predict mis-retrieval. Second, this assumption is also at odds with results from other studies that reported the usual facilitatory mismatch effect in similar structural conditions (i.e., when the intervening element is in a subject position, e.g., Van Dyke and Lewis 2003, Van Dyke 2007, and it is also closer to the retrieval site than the target, Experiments 3 and 4 in Chapter 4).

It is interesting to notice that the off-line effect in comprehension accuracy pointing to encoding interference attested by Jäger et al. (2015) is not isolated. Gordon et al. (2001, 2002) also reported clear off-line mismatch effects in the comprehension of object relative clauses. Similar effects in adults have also been reported in the two self-paced reading experiments of Chapter 4, in which higher comprehension accuracy rates in mismatch conditions as compared to match conditions was consistently found. In addition, findings reported in the acquisition literature systematically attested to higher comprehension accuracy in object relatives when the object and the subject mismatch either in number or gender (Adani 2008; Adani et al. 2010; Adani 2012; Belletti et al. 2012; Adani et al. 2014). Moreover, some of these studies also reported higher comprehension accuracy for subject relatives, where no retrieval is supposedly needed at any time since all the elements are in their canonical position. For instance, Adani (2008)
reported a facilitatory effect of mismatch in subject relatives with Italian children, an effect which had a similar effect size of the one reported in object relatives. Similar results were reported by Adani et al. (2014) with English children with Specific Language Impairment and by Belletti et al. (2012) in Hebrew children (and a numerical difference was also reported for Italian children). Also the two self-paced reading studies presented in Chapter 4 revealed this pattern, with higher comprehension accuracy for mismatch conditions in both subject and object relatives clauses to an extent that was comparable.

Findings from the studies reported above are therefore mixed: two studies report clear evidence for retrieval interference (Van Dyke and McElree 2006; Belletti et al. 2012); eleven studies attest to encoding interference (Gordon et al. 2001, 2002, 2004; Fedorenko et al., 2006; Adani, 2008; Adani et al., 2014; Vasishth and Hofmeister 2014; Kush, Johns, and Van Dyke, 2015; Experiment 1 in Jäger et al., 2015; Experiments 3 and 4 in Chapter 4); five studies attest to interference effects that are compatible both with encoding and retrieval interference (Van Dyke, 2007; Van Dyke and Lewis, 2003; Adani et al., 2010; Van Dyke and McElree 2011; Franck et al., 2015).

Summarizing, I have discussed that an effect observed at the retrieval region is compatible both with encoding and retrieval interference. However, retrieval and encoding interferences differ with respect to the features that are supposed to generate similarity-based interference: while only overlapping features that are targeted by a retrieval cue can generate retrieval, encoding interference arises when two elements share a same feature, regardless of whether or not this feature is also a retrieval cue. Capitalizing on this difference, I manipulated the presence of an agreement feature (gender or number) at the verb region so that the verb either provides or not an agreement cue for retrieval in the comprehension of object relative clauses (ORs) in Italian and English. In Italian, the verb never agrees with the subject in gender, therefore providing no gender cue for retrieval. In English, present tense verbs morphologically express number agreement with the subject (e.g., *criticizes-SG*), thus providing a number cue for retrieval, but past tense verbs do not (e.g., *criticized-Ø*), thus providing no cue for retrieval. Therefore, interference should impact sentence processing differently depending on whether it affects retrieval, encoding or both mechanisms:
(i) if interference affects only retrieval, a facilitatory mismatch effect is expected in the present tense in English, but not in the past tense nor in Italian; 
(ii) if interference affects encoding, a facilitatory mismatch effect is predicted both in Italian and English, whether the verb is in the present or past tense; 
(iii) if interference plays a role both at retrieval and at encoding, the facilitatory mismatch effect should take the form of an interaction with a stronger mismatch effect in English present tense condition (where a retrieval cue is present) than in past tense condition (when no retrieval cue is present), under the assumption that the beneficial effects of mismatch at encoding and retrieval may combine additively.

By investigating agreement dependencies, these experiments also aim at casting light on the rather mix evidence coming from studies investigating the role of agreement features in modulating similarity-based interference, which ranges from studies reporting no effects of agreement features mismatch (e.g., Wagers et al. 2009; Dillon et al. 2013; Tanner et al. 2014; Tucker et al. 2015; Lago et al. 2015) to studies reporting such an effect (Franck et al. 2015; Experiments 3 and 4 in Chapter 4). The lack of a feature mismatch effect that some studies report is inconsistent with the large body of evidence attesting to similarity-based interference when non-agreement features are manipulated (for a review see Engelmann, Jäger, and Vasishth, submitted). This state of affair thus calls for additional empirical evidence in adults and for a theoretical framework accounting for why the beneficial effect usually associated with mismatching features in comprehension is hardly detectable when agreement features are manipulated, which will allow us to cast light on the role of agreement features in the resolution of long-distance dependencies. Here I explore Italian, which has never been investigated so far in adults, and English, for which no effect has been reported to date in grammatical sentences.

2 By mismatch effect I always refer to the feature mismatch between the subject and the object of the relative clause.
5.2 Experiment 1

In Italian, the verb never agrees in gender with the subject, therefore providing no gender cue for retrieval. Therefore, if interference effects arise at encoding, then a gender mismatch effect should be observed, while if interference effects exclusively arise at retrieval, no gender mismatch effect should be observed.

5.2.1 Method

Participants. One hundred and sixty-seven participants took part in the experiment. Participants were all native speakers of Italian (mean age = 33 y.o.) and they were all naïve to the purpose of the experiment.

Materials and design. Thirty-two sets of 4 conditions each were generated in a 2x2 design by manipulating: (i) the gender of the object (masculine vs. feminine), and (ii) the match between the gender of the subject and the gender of object (match vs. mismatch). Noun phrases were always animate and singular. The experimental items consisted in object relative clauses adapted from the sentences of the French experiments reported in Chapter 4 for which semantic reversibility was controlled. All sentences were thus semantically reversible, so that it was not more likely for the agent to perform the action described by the verb than for the patient. In Italian relative clauses, the past participle (sorpreso) never agrees in gender with the subject, therefore remaining in its masculine default form. Examples of experimental items are presented in Table 5.1. Eight lists were created so that each participant was presented with 72 sentences in total, 16 experimental sentences and 56 filler sentences. Experimental sentences were decomposed into 11 regions. Filler sentences were constituted of complex sentences containing movement and/or subordination and subject relatives. They were decomposed in a varying number of reading windows, depending on their length.

---

3 In Italian, past participle agreement holds with (i) unaccusative verb, (ii) passive morphology, (iii) direct object cliticization, (iv) reflexive clitics, and (v) impersonal passive *si* (see Belletti 2016 for a discussion).
Table 5.1. Example of item in the four experimental conditions of Experiment 1.

<table>
<thead>
<tr>
<th>Experimental conditions</th>
<th>Masculine object</th>
<th>Feminine object</th>
</tr>
</thead>
<tbody>
<tr>
<td>Match (MM)</td>
<td>Il/ballerino/che/il/cameriere/ha/sorpreso/beveva/un/cocktail/alcolico. The/dancer-MASC/that/the/waiter-MASC/has/surprised-Ø/drank/a/cocktail/with/alcohol.</td>
<td>La/ballerina/che/la/cameriera/ha/sorpreso/beveva/un/cocktail/alcolico. The/dancer-FEM/that/the/waiter-FEM/has/surprised-Ø/drank/a/cocktail/with/alcohol.</td>
</tr>
<tr>
<td>Mismatch (MF)</td>
<td>Il/ballerino/che/la/cameriera/ha/sorpreso/beveva/un/cocktail/alcolico. The/dancer-MASC/that/the/waiter-FEM/has/surprised-Ø/drank/a/cocktail/with/alcohol.</td>
<td>La/ballerina/che/il/cameriere/ha/sorpreso/beveva/un/cocktail/alcolico. The/dancer-FEM/that/the/waiter-MASC/has/surprised-Ø/drank/a/cocktail/with/alcohol.</td>
</tr>
</tbody>
</table>

Procedure. The experiment was programmed on Ibex Farm, an online experimental javascript-based platform that uses the local machine for timing, thus achieving very accurate timing (www.spellout.net/ibexfarm/docs; Drummond 2013). Sentences were presented on a computer screen in a moving-window self-paced reading paradigm (Just et al. 1982): a series of dashes corresponding to the words of the sentence are presented on the screen, and as soon as the participant presses the space bar the first word appears, replacing the corresponding dashes. Subsequent button press makes the other words appear. Only one word is visible at a given time. Items were presented in a random order.

As soon as the last word of the sentence disappears, a yes/no comprehension question was displayed at the center of the screen and participants were asked to answer the question by clicking with the mouse on one of the two available answers (yes vs. no). Comprehension questions always targeted thematic roles attribution in the relative clause (e.g. Did the waiter surprise the dancer? vs. Did the dancer surprise the waiter?), therefore allowing us to determine if the correct parse of the sentence was built. Instructions encouraged both rapid
reading and correctness in answering the question. The experimental session began with four practice trials. The whole session lasted about 15 min.

### 5.2.2 Results

**Data analyses.** Reading times were analyzed by way of linear mixed-effects regression models, and for question answering data I used generalized linear mixed-effects regression models in the lme4 package (Bates et al., 2015) in R (R Development Core Team, 2016). Only items for which the comprehension question was answered correctly were included in the analysis of reading times. Reading times greater than 3000 ms or less than 100 ms were removed (affecting 2% of the data). No additional outlier removal process was performed. However, in a rapid visual serial presentation task, Staub (2010) showed that the effect of a mismatching intervening subject in object relative clauses was driven by a small set of trials, in particular those trials that have disproportionately long reaction times (see also Lago et al. 2015 for similar results with a self-paced reading task). I thus conducted an additional analysis adopting a more conservative trimming, excluding only reading times exceeding 8000 ms in case the occurrence of an effect depended on inclusion of the right tail of the reading time distribution.

Reading times were log-transformed to normalize residuals and then regressed against two factors that may affect reading times in self-paced reading tasks, namely word length and log list position of the sentence in the stimuli (i.e., longer reading times are associated with longer words and faster reading times with later list position; Hofmeister 2011, Hofmeister and Vasishth 2014). The *residual log reading time* is therefore the dependent variable analyzed here. Error bars in graphs represent standard errors by subject means.

All the predictive factors were dichotomous and centered by coding one level of the factor as -1 and the other as 1. I always used the maximal random-effects structure by participant and by item justified by the data and the random-effects structure always had the same specification as the fixed effects (e.g., if testing for the interaction, the random structure contained the interaction). No correlations between random effects were estimated. These analyses are therefore conservative with respect to the generalizability of the effects of theoretical interest to new participants and items (Barr, Levy, Scheepers, and Tilly 2013). P-values were calculated
by way of the Satterthwaites’s approximation to degrees of freedom with the lmerTest package (Kuznetsova, Brockhoff, and Haubo Bojesen Christensen, 2016).

To assess the gender mismatch effect, analyses on three separate regions were performed: the critical region containing the past participle, the matrix verb region that follows it, and the region containing the second noun phrase (i.e., the subject) susceptible to show an effect of mismatch due to encoding interference (Van Dyke and McElree, 2006).

**Results**

*Comprehension-question accuracy.* Mean accuracy scores of question responses are provided in Table 5.2. Generalized linear mixed effect analysis revealed a significant main effect of gender mismatch ($\beta = -0.366$, $SE = 0.06$, $z = -5.636$, $p < .001$) attesting to higher accuracy scores for mismatch conditions than match conditions. No other effect was significant ($t_s < 1$).

**Table 5.2.** Mean accuracy percentages for comprehension questions by experimental condition in Experiment 1.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender match, Feminine object</td>
<td>75.5</td>
</tr>
<tr>
<td>Gender match, Masculine object</td>
<td>77.9</td>
</tr>
<tr>
<td>Gender mismatch, Feminine object</td>
<td>84.6</td>
</tr>
<tr>
<td>Gender mismatch, Masculine object</td>
<td>83.9</td>
</tr>
</tbody>
</table>

*Reading Times.* The distribution of reading times across the four experimental conditions is reported in Figure 5.1. Non-transformed reading times are plotted for readability, but analyses were conducted on residual log reading times.
Fig. 5.1. Distribution of reading times (in ms) in the four experimental conditions for the different regions of Experiment 1 (correct trials only).

Critical region (past participle). No effect was significant ($t_s < 2$).

Matrix verb. No effect was significant ($t_s < 1$).

Second NP region. No effect was significant ($t_s < 2$).

In line with Staub (2010), I then go through a more conservative trimming, excluding only reading times exceeding 8000 ms. If the effect is on the right skew of the distribution, then it should show up when longer reading times are conserved in the analyses.

Critical region (past participle). Results attested to a significant effect of gender mismatch ($\beta = 0.031$, $SE = 0.014$, $t = 2.253$, $p = 0.025$), with faster reading times for mismatch conditions ($M = 1197$ ms) than match conditions ($M = 1316$ ms). No other effect was significant ($t_s < 1$).

Matrix verb. No effect was significant ($t_s < 1$).

Second NP region. No effect was significant ($t_s < 2$).
5.2.3 Discussion

Experiment 1 found a main effect of gender mismatch in comprehension accuracy, such that gender mismatch conditions were understood better than gender match conditions. However, the effect of gender mismatch only showed up in on-line measures when a more conservative trimming was used, in which case reading times at the critical past participle region were faster in the gender mismatch condition as compared to the gender match condition. This result is in line with results on number agreement by Staub (2010) and Lago et al. (2015), who showed that the number mismatch effect in object relatives lay in the right tail of the distribution, thus being driven only by those trials that appear to be particularly slow, an aspect to which I will return in the General Discussion. It is worth noting, however, that despite manifesting at the integration region, the on-line effect reported in analyses with a more conservative trimming provides evidence for encoding interference, since in Italian the past participle does not agree in gender with the subject, hence gender is not a retrieval cue. As discussed in the Introduction of this Chapter, encoding interference can negatively impact retrieval: feature overwriting, i.e., the mechanism assumed to underlie interference effects at encoding, causes the loss of the overlapping feature in one of the representations and, consequently, a decrease in the activation level of the corresponding memory representation (Oberauer and Kliegl, 2006; Oberauer and Lange, 2008). Since less active elements in memory are more difficult to be accessed, encoding interference can render retrieval more difficult.

It has been argued that encoding effects should manifest immediately at the region in which an element similar to a previously encoded element is encountered (see Van Dyke and McElree, 2006). This prediction is not borne out by results from Experiment 1: no difference between mismatch and match conditions was observed at the subject region. It is important to note, however, that evidence for encoding interference arising at the second noun phrase region is actually scarce in the literature, as discussed in the Introduction: only one study reported clear evidence for such an effect (Kush et al. 2015), while this effect was not interpretable in the studies from Gordon et al. (2001, 2002), due to the confounding provided by frequency and word length, and most of the studies failed to find one (Van Dyke and Lewis 2003; Van Dyke 2007; Van Dyke and McElree 2011; Franck et al. 2015; Jäger et al. 2015), in line with Experiment 1 reported here.
Results from Experiment 1, showing evidence for a facilitatory effect of gender mismatch in Italian object relatives, stand in contrast with results from Belletti et al. (2012), who found no effect of gender mismatch in the comprehension of object relatives in Italian-speaking children. Interestingly, however, Belletti et al. actually found a clear numerical tendency in this direction ($M = 57\%$ vs. $M = 52\%$; $p = .16$ in the ANOVA by subjects and $p = .14$ in the ANOVA by items). Results from Experiment 1 thus suggest that the null result attested by Belletti et al. (2012) may be a Type II error due to lack of power. Finally, the finding attesting to a facilitatory effect of gender match in grammatical sentences aligns with other findings in the adult literature showing facilitatory effects of number match in French (Franck et al. 2015, and Experiments 3 and 4 in Chapter 4). This suggests that the lack of a match effect in grammatical sentences reported in various other studies may also be a Type II error (Wagers et al. 2009, Dillon et al. 2013, Tanner et al. 2014, Lago et al. 2015, Tucker et al. 2015). Moreover, in line with Franck et al. (2015), in Chapter 4 I suggested that the absence of a mismatch effect in these studies may also be due to the systematic presence of ungrammatical sentences in their materials, which may have reduced the reliability of the agreement cues on the verb, therefore reducing its impact on processing.

To summarize, results from Experiment 1 provide support for encoding interference in Italian, by reporting a gender mismatch facilitatory effect when gender is not a retrieval cue. I now turn to Experiment 2, seeking for converging evidence for the role of similarity in agreement features from a different angle, by explicitly contrasting, within the same language (English), the presence vs. absence of an agreement retrieval cue on the verb.

### 5.3 Experiment 2

In English, present tense verbs agree in number with the subject, but past tense verbs do not. If number mismatch only impacts encoding, it should similarly affect the processing of present and past tense verbs. On the contrary, if number on the verb is used as a subject retrieval cue, then two outcomes are possible: if number mismatch impacts retrieval but not encoding, a facilitatory mismatch effect is expected in sentences with verbs in the present tense condition only, since number is not expressed in the past tense verbal morphology. If number mismatch
plays a role both at retrieval and at encoding, a mismatch effect is expected in both present and past tense verbs, but it should be greater in the former which carries a number retrieval cue.

5.3.1 Method

Participants. One hundred and thirty participants took part in the experiment. They were all students at the University of Connecticut and they received course credit for their participation. They were all native English speakers and naïve about the purpose of the experiment.

Materials and design. Thirty-two sets of four conditions each were generated in a 2x2 fully crossed design manipulating: (i) the match between the number of the subject, which was always singular, and the number of the object, which could be either singular or plural (match vs. mismatch), and (ii) the presence of an agreement cue on the verb (cue vs. no cue). The presence of an agreement cue on the relative verb was manipulated by taking either present tense verbs (e.g., criticizes-SG), which exhibit subject-verb number agreement, or past test verbs, in which agreement is not morphologically expressed on the verb (e.g., criticized-Ø). Sentences were all object relative clauses with animate subjects and objects. As for Experiment 1, the experimental sentences were an adaptation of the experimental sentences used in Experiments 3 and 4 of Chapter 4, for which thematic roles reversibility was controlled for. Therefore, no semantic cue was available to assign thematic roles. Examples of experimental items are presented in Table 5.3. Verbal agreement was manipulated on the verb of the relative clause, while the matrix verb was kept in the past form in order to restrict the agreement cues present in the sentence. In order to control for potential spillover effects, which manifest when the reading times measured in region \( n \) are influenced by reading times in region \( n-1 \), an adverb was introduced before the critical region of interest (i.e., the relative verb). The critical verb (criticize-s/-d) was always followed by a complex quantifier phrase (e.g., most of) with a temporal modifier as complement followed by the matrix verb. Eight lists were created so that each participant was presented with 72 sentences in total: 16 experimental sentences and 56 filler sentences. As for Experiment 1, filler sentences consisted of complex sentences with movement and/or subordination and subject relatives. Experimental sentences were decomposed into a number of reading windows varying between 15 and 17, each containing
either a content word or a grammatical word. Filler sentences were decomposed in a varying number of reading windows, depending on their length.

Table 5.3. Example of item in the four experimental conditions of Experiment 2.

<table>
<thead>
<tr>
<th>Experimental conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agreement cue</td>
</tr>
<tr>
<td>Match (SS)</td>
</tr>
<tr>
<td>Mismatch (PS)</td>
</tr>
<tr>
<td>No Agreement cue</td>
</tr>
<tr>
<td>Match (SS)</td>
</tr>
<tr>
<td>Mismatch (PS)</td>
</tr>
</tbody>
</table>

Procedure. I used a noncumulative self-paced paradigm (Just et al. 1982) where participants read sentences in a word-by-word fashion by pressing the space bar to display the next word. The experiment was programmed with the E-prime software (Schneider, Eschman, and Zuccolotto, 2012). Each trial began with a fixation cross (400 ms) followed by an interstimulus blank screen (150 ms) and then the word-by-word presentation of the sentence. Each trial was separated from the other by an instruction in which participants were asked to press the space bar as soon as they were ready to continue. After each sentence a yes/no comprehension question was presented on the computer screen. An interstimulus blank screen (150 ms) separated each sentence from the corresponding comprehension question, which appeared at the center of the screen. Comprehension questions specifically targeted the critical relative verb to determine whether the correct parse was built (e.g., Did the waiter criticize the dancer? vs. Did the dancer criticize the waiter?).
Instructions encouraged both rapid reading and correctness in answering the question. Items were presented in a fixed pseudo-random order constrained such that no more than two consecutive trials were experimental sentences. Each experimental session began with four practice trials. Three breaks of 1-minute each were administrated during the task. The whole session lasted about 20 minutes.

5.3.2 Results

Data analyses. The same data analyses conducted for Experiment 1 were used here. To assess the number mismatch effect and its interaction with the presence of a number cue at the verb, I performed analyses on three separate regions: the critical region containing the embedded verb (criticize-s/d), the region immediately following the past participle to investigate potential spillover effects, and the region containing the second noun phrase (i.e., the subject) susceptible to show an effect of mismatch due to encoding interference (Van Dyke and McElree 2006).

The error rate in comprehension accuracy for object relatives was particularly high (30% incorrect responses). However, the near ceiling performance to comprehension questions for filler items (95% correct responses) suggests that the high error rate in object relatives was not due to a general lack of attention during the task, but rather reflects a genuine difficulty in the processing of experimental sentences. Yet, since I was interested in investigating the effectiveness of number cues in driving structure building, I restricted the analyses to items for which participants built the correct parse thus leading to a correct response.

Results

Comprehension question accuracy. Mean accuracy scores of question responses are provided in Table 5.4. Generalized linear mixed effect analysis revealed a marginal effect of number mismatch ($\beta = -0.088$, $SE = 0.051$, $z = -1.713$, $p = .086$), attesting to numerically higher comprehension accuracy for number mismatch than number match conditions. No other effect was significant ($t < 1$).
Table 5.4. Mean accuracy scores of question responses in percentage by experimental condition in Experiment 2.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number match, Agreement cue</td>
<td>64.1</td>
</tr>
<tr>
<td>Number match, No agreement cue</td>
<td>65</td>
</tr>
<tr>
<td>Number mismatch, Agreement cue</td>
<td>66.5</td>
</tr>
<tr>
<td>Number mismatch, No agreement cue</td>
<td>68.6</td>
</tr>
</tbody>
</table>

Reading Times. The distribution of reading times across the four experimental conditions in correct trials is reported in Figure 5.2.
Fig. 5.2. Distribution of reading times (in ms) in the four experimental conditions for the different regions of Experiment 2 for correct trials only.
Critical region. No effect was significant \((t_s < 1)\).

Spillover region. No effect was significant \((t_s < 2)\).

Second NP region. No effect was significant \((t_s < 2)\).

As for Experiment 1, in order to investigate whether the effect was in the right tail of the distribution, I conducted an additional analysis with a more conservative trimming in which only reading times longer than 8000 ms were eliminated.

Critical region. No effect was significant \((t_s < 2)\).

Spillover region. A marginal significant interaction between the number match and the presence of an agreement cue on the verb was found \((\beta = -0.026, SE = 0.015, t = -1.888, p = .059)\), attesting to a marginally significant effect of number mismatch in the agreement cue condition (i.e., present tense verbs) \((\beta = -0.072, SE = 0.040, t = 1.794, p = .073)\) revealing faster reading times at the spillover region in the number mismatch condition \((M = 488 \text{ ms})\) than in the number match condition \((M = 562 \text{ ms})\), but no effect was observed in the no agreement cue condition \((t < 1)\) (i.e., past tense verbs).

Second NP region. No effect was significant \((t_s < 1)\).

5.3.3 Discussion

Results from Experiment 2 revealed a tendency for higher comprehension accuracy for mismatch than match conditions, an effect that did not interact with the presence of a number retrieval cue on the verb. The lack of interaction is in line with the hypothesis that similarity in number features affects encoding only, since the effect of feature mismatch is not modulated by the presence of a number retrieval cue at the verb. This conclusion is in line with results from Experiment 1, although here the potential role of similarity at retrieval can be assessed, which does not show up in off-line results. The fact that, unlike Experiment 1, the accuracy mismatch effect in Experiment 2 is only marginally significant is possibly due to the global low comprehension accuracy rate for experimental sentences, which was much lower as compared to the accuracy rate in Experiment 1 (80% of correct responses for experimental sentences in Experiment 1 vs. 66% in Experiment 2), which may therefore have reduced the chances to
observe a fully significant effect of number mismatch. For the current purpose, however, it is sufficient to observe that the observed tendency goes in the expected direction and that no interaction was observed between conditions with and without a retrieval cue on the verb, therefore pointing towards a mismatch effect arising at encoding.

As in Experiment 1, an on-line effect of feature mismatch was only observed when a more conservative trimming was used. However, unlike Experiment 1, no online effect of feature mismatch was observed at the critical verb region, but the effect was found at the region immediately following it. This effect attested to a tendency for a number mismatch effect only when an agreement cue is present on the verb. Since the mismatch effect was modulated by the presence of a retrieval cue, this result provides some indication for an effect of mismatch arising at retrieval. However, the absence of a facilitatory mismatch effect in the past tense condition is at odds with the Italian results that showed a significant online mismatch effect in the absence of a retrieval cue on the verb. The absence of a mismatch effect in the English past tense condition may be due to a lack of power due to the smaller sample size of this experiment as compared to the Italian one. I will discuss the implications of this hypothesis in the General Discussion.

All in all, results from Experiment 2 provide evidence for a facilitatory number mismatch effect in grammatical sentences, extending to English results observed in French (Franck et al., 2015, Experiment 1; Experiments 3 and 4 in Chapter 4) and Italian (Experiment 1).

5.4 General Discussion

5.4.1 Summary of the findings

I started by reviewing several studies investigating interference effects in sentence comprehension and I pointed out that, despite the rather common assumption that the main operation at play in the resolution of long-distance dependencies is the retrieval of the long-distant element (e.g., McElree 2000, 2006; McElree et al. 2003; Lewis and Vasishth 2005), only results from two studies report clear evidence attesting to retrieval interference (McElree and Van Dyke 2006; Belletti et al. 2012), while results from several studies that have been interpreted in terms of retrieval interference are actually also compatible with interference
arising at encoding (e.g., Van Dyke 2007; Van Dyke and Lewis 2003; Van Dyke and McElree 2011; Franck et al. 2015), and several studies provide direct evidence for encoding interference (Gordon et al. 2001, 2002, 2004; Fedorenko et al. 2006; Vasishth and Hofmeister 2014; Kush, Johns, and Van Dyke, 2015; Jäger et al., 2015 Experiment 1; Adani, 2008; Adani et al. 2014; Experiments 3 and 4 in Chapter 4). This state of affair is at odds with current models of sentence comprehension granting little or no role to encoding mechanisms in sentence comprehension. For instance, the ACT-R model (Lewis and Vasishth 2005; Engelmann et al. 2015) does not include a mechanism to account for encoding interference. The picture is further complicated by studies showing that interference effects are not always detected, even in the presence of a retrieval cue at the verb. In particular, studies investigating the role of agreement features in the comprehension of grammatical sentences are mixed, some attesting to a facilitatory effect of mismatch (Franck et al. 2015; Experiments 3 and 4 in Chapter 4), others attesting to a facilitatory effect of match (Jäger et al. 2015) or to a lack of effect (Wagers et al. 2009; Dillon et al. 2013; Tanner et al. 2014; Tucker et al. 2015; Lago et al. 2015). This latter set of studies systematically reported a facilitatory effect of agreement feature mismatch in ungrammatical sentences (an illusion of grammaticality), but no corresponding effect in grammatical sentences (i.e., no illusion of ungrammaticality).

With the aim of providing a solid basis for theoretical elaboration, I conducted two self-paced reading experiments specifically designed to collect further evidence for the role of agreement feature mismatch in the comprehension of grammatical sentences and to consider what implications such findings would have for cue-based retrieval models of sentence processing. More specifically, the aim was to investigate whether mismatch effects depend on the presence of agreement cues on the verb or not, in order to determine whether interference arises at retrieval, encoding or both. I tested gender and number subject-verb agreement respectively in Italian and English object relative clauses, each language providing a different test case. In Italian, the verb never agrees in gender with its subject, therefore providing no cue for subject retrieval. In English, the verb is morphologically marked for number agreement in the present tense, thus providing a subject retrieval cue, but not in the past tense. If a mismatch in agreement features affects encoding interference only, then a facilitatory mismatch effect is expected independently of whether the verb carries an agreement cue and should therefore be found in both Italian and English (at the present and past tenses). If feature mismatch affects retrieval interference only, then a facilitatory effect of mismatch should emerge in the English present.
tense condition only. Finally, if both encoding and retrieval interference are at play, a facilitatory mismatch effect should show up in Italian as well as in both conditions in English, but it should be stronger in the present tense than in the past tense.

In line with the hypothesis that agreement feature mismatch affects encoding, results showed that: (a) feature match in the absence of agreement cue on the verb has a weak effect online, with longer processing at the verb in match than mismatch conditions, an effect that was carried mainly by long (over 3000 ms) trials (I take the fact that this effect was present in Italian and absent in English as evidence for its weakness); (b) feature match has a more robust effect offline, with feature match causing more errors of comprehension than feature mismatch (significant effect in Italian, marginal effect in English); (c) the online effect of feature match in English exhibits a marginal interaction with the presence of a cue on the verb: match is slower than mismatch when the verb is marked for agreement.

All in all, results in Italian and English, combined with studies from Franck et al. (2015) in French and Experiments 3 and 4 in Chapter 4, show that similarity in agreement features does modulate the comprehension of grammatical sentences and does so even in the absence of an agreement cue on the verb. This provides additional support to the hypothesis, already put forward in Chapter 4, that the absence of such effects in some studies (e.g., Wagers et al. 2009; Dillon et al. 2013; Lago et al. 2015) may be due to the fact that these studies systematically included ungrammatical sentences in their materials, whereas studies that showed significant effects all involved only grammatical sentences. It is possible that the introduction of a wide portion of ungrammatical sentences prompts participants to develop artificial strategies, which hide the effect (Franck et al. 2015). Thus, I will consider the fact that similarity in agreement features influences sentence comprehension as an established, cross-linguistic finding, and discuss the possible mechanisms underlying the effects reported. I will first discuss to what extent a cue-based retrieval memory model such ACT-R (Lewis and Vasishth 2005) can account for these results, given the prominence of this model in the literature on interference. I will then sketch how a different approach to parsing, based on self-organization (Kempen & Vosse, 1989; Tabor & Hutchins, 2004; van der Velde & de Kamps 2006; Vosse & Kempen, 2000), implements retrieval and encoding interference, and argue that it naturally captures the data at hand.
5.4.2 An encoding interference mechanism in ACT-R: Activation leveling

The results reported in this Chapter provide evidence that encoding interference is at play in sentence comprehension, since interference effects were found in the absence of a retrieval cue on the verb. In the introduction, I noted that prior researchers suggested that one possible mechanism underlying encoding interference is feature overwriting (Nairne 1990; Oberauer and Kliegl 2006). In line with this view, Vasishth, Jäger and Nicenboim (2017) argued that a statistical model implementing feature overwriting provides a better fit to account for the difficulty associated with a match in agreement features than a statistical model implementing a cue-based retrieval account. However, feature overwriting alone does not specify how parsing works, so I now take up the question of how this mechanism could be implemented in a parsing model, like ACT-R, which is the most highly developed model of interference effects in sentence processing. As a reminder, feature overwriting is a mechanism in which, at the point of encoding, if two arguments share a feature, then one of the two ends up losing the feature: in one version, it is the first encoded element that stochastically loses the feature (Nairne 1990); in another, either the first or the second element may lose it (Oberauer and Kliegl 2006); one might also suppose that just the second element loses the feature. However, although such mechanism causes featural changes at encoding, if implemented in ACT-R, it would have no effect at retrieval whenever the overlapping feature is not a retrieval cue, which is the case for Italian and English past tense here as well as for several studies in the literature (e.g. Gordon et al. 2001, 2002; Hofmeister and Vasishth 2014), because cue-based retrieval is only sensitive to features that are specified at the integration region. I therefore consider another way of interpreting feature overwriting which fits naturally into the framework of ACT-R.

Feature overwriting is based on the principle that elements that share features compete for them; generally, competition effects in ACT-R are treated as competition for activation (see Lewis & Vasishth 2005 on the fan effect). Thus, it is natural to assume that featural similarity at encoding produces activation competition. The activation of chunks in memory in ACT-R is determined by: (i) the base activation level of the chunk, which is determined by the past activations and reactivations of the chunk, and by the time elapsed since the last reactivation because of decay, (ii) the strength of the association between each retrieval cue and the chunk, that is the uniqueness with which the cue identifies the chunk (fan effect) and (iii) a random noise
component. To these conditions, I add the further assumption (iv) that when a new chunk shares a feature with one chunk that has already been encoded, the activations of all chunks sharing that feature become more equal. I call this the leveling effect. Thus, if two elements are not already equal in activation, the higher one goes down and the lower one goes up. Under the fan mechanism (Anderson, 1974; Lewis & Vasishth, 2005), a fixed amount of activation associated with a retrieval cue is distributed over all chunks in memory that share that feature. Since a chunk has to reach a fixed, high threshold of activation in order to be selected as a target of retrieval, conditions in which a feature is spread across two or more chunks (match) have slower retrieval than conditions in which the feature is unique to a chunk (mismatch). The current proposal extends this principle of activation-sharing to all features that are shared across chunks, whether they are retrieval cues or not. Since the levelling mechanism applies to all features, including features that are not retrieval cues, one might consider eliminating the fan mechanism in favor of levelling (thus treating what have been viewed as retrieval interference effects as encoding interference effects). However, the English online results suggest (weakly) that retrieval cues at the verb enhance the effect of agreement feature similarity; to generate this enhancement, I thus keep the fan mechanism in the model (which is, in any case, independently motivated by data from Van Dyke and McElree 2006 and Belletti et al. 2012 and possibly by data stemming from those studies reviewed above which are compatible with both encoding and retrieval interference if future research will determine that the source of the effect is at retrieval). Finally, I make the assumption that, on average, in grammatical sentences, putting aside effects of leveling, subjects have more activation than other chunks (Franck and Wagers 2015). Therefore, leveling always has the effect of reducing the subject’s activation and increasing the activations of other chunks.

This model generates most of the effects reported in the present paper. First, the model generates the on-line agreement feature match effect (i.e., match slower than mismatch) when no agreement cue was present at the verb by lowering the activation of the candidate leading the activation race (i.e., the subject), so the race takes longer to conclude. Second, the model generates the agreement feature match effect in off-line comprehension questions because reducing the level of activation of the subject reduces the chances that it is correctly selected, and thus that the sentence structure is correctly built. Moreover, as I anticipated above, the model generates the marginal online interaction between number match and the presence of an
agreement cue on the verb because, in the present tense condition, fan favors the mismatch over the match condition, over and above the effect of leveling.

In this version of ACT-R that I propose, encoding interference effects are due to leveling whereas retrieval interference effects are due to fan. Leveling causes any two chunks that are similar to equalize their activation and this will result in confusion among them if there is a retrieval process that encompasses both of them. This amounts to an encoding interference mechanism, as it affects the representations in virtue of the way they are encoded. The fan mechanism is invoked in ACT-R when a cue-based retrieval occurs, and it only applies to features that are cued at the verb. So fan is specifically a retrieval interference mechanism.

One aspect of the data that the model does not generate is the more robust appearance of encoding interference effects offline than online (both in the two Experiments presentend here and in Experiments 3 and 4 in Chapter 4). Although this contrast may occur for uninteresting reasons (e.g., the fact that two different measurement methods are used online and offline), we propose another model in the following section that naturally accounts for it, in addition to the other effects reported.

5.4.3 A Self-Organized Sentence Processing (SOSP) account of encoding and retrieval interference effects

Cue-based retrieval in ACT-R is a mechanism specifically concerned with relating elements in long-distance dependencies. Self-organized sentence processing (SOSP; see Kempen & Vosse, 1989; Tabor & Hutchins, 2004; van der Velde & de Kamps 2006; Vosse & Kempen, 2000) generalizes the use of cue-based retrieval to all structure-formation. I argue below that this choice provides a simple and well-motivated way of capturing not only the encoding and retrieval effects observed, but also their (possible) interaction with the online vs. offline distinction.

Under SOSP, each perception of a word activates a treelet (a mother node with a finite set of daughter nodes) in memory similar to a chunk in ACT-R (see also Fodor 1998a, 1998b, 2017; Marcus, 2001). As activated treelets accumulate in memory, they attempt to combine in all possible ways with other activated treelets, subject to the restriction that daughter nodes only attempt to link with mother nodes of other treelets (i.e., no daughter-to-daughter or mother-to-
mother connections and no within-treelet connections). Each mother and each daughter is a vector of semantic and syntactic features encoding the properties of the attachment site; treelets with multiple daughters make the daughters available for attachment in series, reflecting word-order constraints. All links start out, at the beginning of processing, with strength 0, and the strengths are constrained to lie in the interval $[0, 1]$. Each link strength grows and/or shrinks over time following two principles: (a) the better the match between the feature vectors at its ends, the more rapidly the link strength grows, and (b) links for the same attachment site compete in a winner-take-all fashion. Furthermore, the numerical values are noisy so there is some variation in how precisely the system adheres to these two principles. Finally, within the constraints imposed by the semantic and grammatical requirements of a treelet, the feature vectors on opposite ends of a link migrate toward the same values as the link grows stronger (a bidirectional feature passing mechanism). In self-paced reading and language production, the model moves to the next word when the attachments for the current word have nearly stabilized. After a sentence has been processed, the link strengths return to 0, but the treelets linger in the states they have gravitated to under the bidirectional feature passing. This supports immediate reconstruction of the tree for purposes of repeating the sentence or answering questions about it.

To illustrate how this system works, I will show how it generates agreement attraction effects in a case related to the cases reported here. In the standard preamble continuation paradigm of Bock & Miller (1991), Barker et al. (2001) found that participants were more likely to produce a plural verb in examples like (1a), where the subject and the attractor share a fine-grained semantic feature (boat- hood) than (1b), where they do not.

\begin{tabular}{l}
(1a) The canoe near the sailboats…
(1b) The canoe near the cabins…
\end{tabular}

In SOSP, when the first two words of the preamble have been spoken, a treelet is activated by canoe that has features on its mother including $[+\text{Noun}, +\text{Boat}, +\text{Singular}]$. This treelet begins to combine with a verbal treelet (whose lexical anchor has not yet been generated) and causes the subject daughter of the verb, which is already specified to be $[+\text{Noun}]$, to gravitate toward a feature vector that includes $[+\text{Noun}, +\text{Boat}, +\text{Singular}]$ (the gravitation is a gradual process in which features originally set to the value 0 continuously increase their values toward a
maximum of 1, driven by the growing strength of the link between the canoe treelet and the subject daughter of the verb). While this is happening, near the sailboats is spoken and another treelet is generated by sailboats, bearing the features [+Noun, +Boat, +Plural]. Because of the growing presence of the feature [+Boat] on the verb treelet, there is an enhanced tendency for the sailboats treelet to form a link with the subject daughter of the verb because sailboats also has the feature [+Boat]. This possibility of erroneously attaching the sailboats treelet as the subject of the verb follows from the assumption that all possible combinations attempt to form, in virtue of principles (a) and (b) above. During its interaction with the subject site, sailboats can cause the number feature of the subject daughter to gravitate to [+Plural]. This can happen even though canoe ultimately wins the competition to attach as the subject of the verb. When it is time to generate the verb, the model produces the verb that is most highly activated within the verb treelet. Consequently, in Barker et al.’s design, there is a greater chance that a plural verb will be produced after the canoe near the sailboats than after the canoe near the cabins, since in the latter case, cabins does not have the feature [+Boat] and therefore its treelet is less prone to being momentarily attached as the subject of the verb.

Turning now to the present experiments, the model is consistent with the results. When the embedded verb is processed, the model makes an attachment site available for its subject and one for its object. As the attachment between the embedded subject and the subject slot on the verb begins to form, its features, including its agreement features, are gradually transferred to the subject slot on the verb (as in the Barker et al. example above), while a similar mechanism takes place as the extracted object starts to attach to the object slot. Since all treelets compete for attachment at all sites, there is subject-object competition: the extracted object also attempts to attach as the subject of the embedded verb and the embedded subject attempts to attach as the object of the embedded verb. The intensity of this competition is stronger if the embedded subject and extracted object are featurally more similar (i.e., match condition), because this makes each of them fit the other’s slots on the verb better. Since the processor waits to continue until this attachment decision has been relatively resolved, the match condition produces longer processing times than the mismatch condition. However, this effect is weak at the beginning because the features have only been weakly transferred to the verb treelet. Over time, the effect of the feature transfer will become stronger and, as a result, the effect of match will become stronger. The dynamics of feature transfer is thus responsible for two aspects of the effects of match reported in the two experiments reported in this Chapter: the fact that the on-line effect
is only statistically reliable when slow trials are taken into account, and the fact that the effect becomes even stronger when a later measure is taken, as is the case of off-line responses. The model also explains the finding from Experiment 2 that feature match has a stronger role when the verb is marked for number: if the verb is marked for agreement, in the mismatch condition the subject slot is the only slot endowed with an agreement feature that matches that of the subject, thus strengthening the correct link between the subject and the verb, in contrast to the match condition in which the feature on the subject slot of the verb also matches that of the object.

A key property of the self-organization approach is that it does not distinguish between encoding and retrieval mechanisms, unlike the extended version of ACT-R proposed above, which has one mechanism accounting for retrieval interference (fan) and another mechanism accounting for encoding interference (activation leveling). Rather, the effect of match in SOSP arises in two different structural relations: (1) in the relation between the two noun phrases: when the noun phrases are encountered and have to be attached to their slots, the similarity in their agreement features plays a role in attracting each to the other’s slot; (2) in the relation between the noun phrases and the verb: when the verb is encountered, its agreement feature attracts noun phrases that have the same agreement feature to its subject slot. One may therefore wonder why the effect of verb marking (the stronger effect of match found for present than past verbs) appeared on-line but not off-line. This observation receives a natural explanation in the model, under the view that what matters off-line is which features have gotten onto the verb, and not whether they have been originally been encoded on the verb due to morphology or due to being transferred during the on-line processing.

---

4 This idea is similar in spirit to Ferreira’s hypothesis that the parser may first engage in a shallow processing of the sentence, possibly because this process is more efficient and less resource-consuming, and would only stabilize on the structure once a question is asked and requires that a decision be taken (e.g., Ferreira 2003).
5.5 Conclusions

In two self-paced reading experiments conducted in Italian and English, strong evidence for encoding interference have been observed, while only weak evidence for retrieval interference was observed. I proposed to implement an encoding mechanism in ACT-R, which in its current formulation does not predict encoding interference effects, by adding a mechanism of activation leveling. Equipped with this additional mechanism, ACT-R successfully accounts for both encoding and retrieval interference. Then, I showed how SOSP can also account for the results at hand. However, whereas ACT-R maintains the well-accepted distinction between encoding and retrieval interference by assuming that two different mechanisms underlie those effects (i.e., fan and leveling), under SOSP the distinction between encoding and retrieval interference vanished, since a single mechanism underlies both effects (i.e., a mechanism of attraction among elements in the process of structure building). It seems that SOSP has the advantage to be simpler than ACT-R, not only because it needs only one mechanism to account for both encoding and retrieval interference rather than two, but also because in SOSP the interference effects arise directly from the parsing mechanism itself, which is not the case in ACT-R. Whether or not retrieval and encoding interference should be ultimately reduced to a single mechanism remains an open question for future research.
Chapter 6

Conclusions

My research aimed to combine the knowledge accumulated in syntactic theories with that accumulated in cognitive science in order to better understand how formal properties and structural operations interact with the memory operations during sentence comprehension. Ultimately, my thesis was concerned with the relationship between the grammar and the parser. One privileged ground to investigate this question is represented by long-distance dependencies, which are both constrained by the grammar and the functioning of the memory system. I proposed to integrate two theoretical frameworks, Featural Relativized Minimality, a pure syntactic theory, and the Cue-based memory model, a processing model, with the aim to attain a higher explanatory adequacy concerning how formal properties interact with memory operations in sentence processing.

I argued that only a subset of syntactic features is responsible for generating grammatical violations when shared both by the target and the intervener, while other linguistic features modulate the memory processes at play in the resolution of long-distance dependencies without generating a shift in the grammaticality of the sentence. I suggested that the first family of features is tied to constraints lying in the grammar, while the second family is tied to processing constraints that ultimately lie in the functioning of our memory system. I then systematically analyzed what are the memory operations underlying the processing of long-distance dependencies. With this aim, I focused on the second family of features, those that are supposed to be tied to the workings of the memory system, by enlarging my investigation to agreement features such as gender and number. If these features modulate the memory processes at play in the resolution of long-distance dependencies, a dissimilarity in terms of these features between the target and the intervener should facilitate the processing of the sentence, due to reduced similarity-based interference at retrieval, encoding or both. I argued that dissimilarity in terms of agreement features plays a major role at encoding, even though it is generally assumed that one of the major operation that the parser undertakes in the resolution of long-distance dependencies is the retrieval of the long-distance element. Finally, I investigated
whether production and comprehension are both sensitive to similarity-based interference, and I suggested that this is the case, even though I argued that these effects originate from different processes in the two cases, respectively the process of controller selection in agreement production and the process of structure building in the comprehension of complex structures. In what follows, I first summarize the main findings of the thesis, and then I develop further thoughts for future research.

6.1 Summary of the most relevant findings

6.1.1 Intervention effects in wh-islands

Experiments 1-3 from Chapter 2 provide a first systematic empirical investigation of intervention effects in wh-islands as a function of the similarity between the target and the intervening element. In three acceptability judgment studies, I manipulated the lexical restriction of the extractee and the intervener giving rise to four configurations: Bare Identity (e.g., *What do you wonder who built?*), Inverse Inclusion (e.g., *What do you wonder which engineer built?*), Inclusion (e.g., *Which building do you wonder who built?*), and Complex Identity (e.g., *Which building do you wonder which engineer built?*). Predictions from fRM are globally borne out, with conditions with a higher degree of featural similarity between the extractee and the intervener (i.e., Bare Identity and Inverse Inclusion) being less acceptable than conditions with a lower degree of similarity (i.e., Inclusion and Complex Identity). However, Complex Identity was systematically found to be the highest rated condition, a finding that is problematic for fRM, since the theory predicts Inclusion and Complex Identity to be rated on a par (see Atkinson et al. 2015 for similar results in English). In Chapter 2, I proposed that, while the increased acceptability of Inclusion as compared to Bare Identity and Inverse Inclusion is accounted for in terms of the set-theoretic approach of fRM, the higher acceptability of Complex Identity as compared to Inclusion is due to extra-grammatical factors and in particular to the richer semantic information carried by the lexical restriction, which contributes in distinguishing elements in memory, therefore reducing similarity-based interference (see also Hofmeister et al. 2007, 2013). However, this solution did not appear to be very parsimonious, as it amounted to say that a very same factor, i.e., lexical restriction,
sometimes generates a fRM-type effect and sometimes generates an effect that is external to fRM.

In Chapter 3, I was thus concerned with disentangling those features that are responsible for fRM-type effects from features that have effects that are external to fRM. In 6 acceptability judgment experiments, I showed that semantic features such as animacy and reversibility of thematic roles also modulate the acceptability of wh-islands, although mildly. The amelioration ensured by dissimilarity in terms of these features was nevertheless fully comparable with that ensured by lexical restriction (i.e., same effect size). Moreover, lexical restriction and semantic features modulate the acceptability of both wh-islands and fully grammatical long-distance dependencies (e.g., extraction out of a declarative). Under the assumption that principles of the grammar define the boundaries of grammaticality, this result suggests that the amelioration associated with the increasing distinctiveness between elements in memory ensured by these features is extra-grammatical, as it is observed to a similar extent in both grammatical and ungrammatical sentences. In addition, while the effect of lexical restriction, animacy and reversibility of thematic roles appeared to be mild, the effect of syntactic features involved in the definition of the landing site of movement (+Q) was much stronger. From these observations, I suggested to distinguish between two families of features: those that when shared both by the target and the intervener are responsible for setting the boundaries of grammaticality, which fRM is concerned with and which are exclusively features triggering movement, and those that when shared both by the target and the intervener are responsible for the modulation observed among structures belonging to the same grammatical side, whose effect is external to fRM and tied to the functioning of the memory system.

6.1.2 Mechanisms underlying the comprehension and production of grammatical long-distance dependencies

Capitalizing on conclusions from Chapter 3, the first aim of Chapter 4 was to investigate what are the memory processes underlying similarity effects tied to the functioning of the memory system, and in particular whether these effects are observed because of similarity-based interference effects arising at encoding, retrieval or both. I focused on agreement features for two main reasons: first, under the hypothesis put forward in Chapter 3, similarity effects in
terms of these features are expected to lead to processing difficulty, without undermining the grammaticality of the sentence; secondly, results from studies testing for similarity in agreement features sometimes report no effect of these features in modulating similarity-based interference in grammatical sentences, therefore calling into question the hypothesis that these features significantly, even if mildly, modulate processing difficulties.

The second aim of Chapter 4 was to investigate whether both production and comprehension are sensitive to similarity-bases interference effects, capitalizing on theoretical approaches suggesting that similar mechanisms underlie both comprehension and production.

In two experiments using the two force-choice response time paradigm (Experiments 1 and 2), I investigated attraction effects due to the intervening subject in object-past participle number and gender agreement in French, a relatively understudied agreement configuration. Results replicate classical findings attesting to object attraction in subject-verb agreement and extend them to subject attraction in object-verb agreement. I interpreted these findings in light of the Feature and Controller Selection model (Franck, in press), according to which attraction effects are the result of similarity-based interference from the intervening element (the subject in this case) in the process of selecting the agreement controller (the object in this case).

In two self-paced reading experiments in French (Experiments 3 and 4), I tested the same configurations in comprehension. Results showed that features such as gender and number modulate the ease with which object relative clauses are understood: comprehension accuracy was higher in mismatch than match conditions, and the mismatch effect was sometimes visible online at the critical integration region (i.e., the past participle, Experiment 4), an effect that I interpreted at first in terms of similarity-based interference during the retrieval of the long-distance element. However, results from these experiments also reported a facilitatory mismatch effect in the comprehension of sentences not containing a long-distance dependency, such as subject relative clauses. Since no retrieval is needed in subject relatives at any time, the mismatch effect observed in these structures is likely to reflect encoding interference. However, since the mismatch effect was comparable in subject and object relatives, and in subject relatives no retrieval is needed at any time since both arguments are in their canonical position, I suggested to interpret the mismatch effect observed in object relatives as also resulting from encoding interference.

Taken together, results from the four experiments of Chapter 4 suggest that both production and comprehension are sensitive to similarity-based interference. I suggested that, on the one
hand, mismatch effects in comprehension reflect similarity-based interference in the process of structure building, such that an intervening element matching the agreement feature of the target generates more interference than an intervening element mismatching it. On the other hand, mismatch effects attested in production reflect similarity-based interference during the process of controller selection in agreement computation, such that an element with controller-like features generates substantial interference during controller selection. On this view, controller selection is actually independent of agreement feature mismatch per se, the detrimental effect of agreement feature mismatch exclusively resulting from the fact that this is the only condition in which controller selection errors are observable, therefore explaining why the mismatch effect is flipped in the comprehension and production of complex sentences involving a long-distance dependency.

The conclusion from Chapter 4 that agreement features played a major role at encoding rather than at retrieval was further investigated in Chapter 5, in which I systematically tested the source of similarity-based interference effects in object relatives by comparing conditions in which an agreement retrieval cue was present at the retrieval region (i.e., the verb) and conditions in which no retrieval cue was present, by taking advantage from properties of Italian and English. Comprehension accuracy results from both languages consistently reported a facilitatory mismatch effect in accuracy measures of sentence comprehension, regardless of the presence of a retrieval cue on the verb, therefore providing evidence for encoding interference. On-line mismatch effects were attested both in Italian and English, but in English the effect was attested exclusively when a retrieval cue was present on the verb, thus providing some support for retrieval interference. However, on-line effects in both languages were exclusively driven by particularly slow trials. In order to account for these findings that point to the key role of encoding during sentence processing, I discussed how encoding interference can be generated by a model such as ACT-R (Lewis and Vasishth 2005) and by a self-organized parser (SOSP, e.g., Tabor and Hutchins 2004). I discussed that a new mechanism should be introduced in ACT-R to generate encoding interference (activation leveling). Equipped with this additional mechanism, ACT-R appears to successfully account for both retrieval and encoding interference. Then I discussed how SOSP accounts for interference effects and I show that also this model successfully accounts for both encoding and retrieval interference. However, while ACT-R assumes two different mechanism, one for retrieval interference (fan) and another for
encoding interference (leveling), SOSP accounts for both types of interference with a single structure building mechanism of attraction among elements during structure building.

6.2 Further thoughts

The findings reported in this thesis provide strong evidence that in the presence of an intervening element that is similar in its featural specification to the target element, the parser experiences substantial difficulties in processing the structure. However, an important question concerns the metric of features that enter in the calculation of the similarity: are all linguistic features weighted the same in generating intervention/similarity-based interference effects or are similarity on some features more disruptive than similarity on other features? The findings I provided show that core syntactic features playing a key role in the syntactic derivation of the structure (e.g., +Q) are weighted more strongly in the calculation of similarity than other syntactic and semantic features (e.g., lexical restriction, animacy, reversibility of thematic roles, number, gender). In particular, only the former can cause derivation failure when shared both by the target and the intervener, while similarity in terms of the latter significantly, even if mildly, modulates the acceptability of the sentence, without ultimately undermining its grammatical status. I thus suggested that the effect of similarity in terms of this latter family of features is tied to the functioning of our memory system, while the effect of similarity in terms of features belonging to the first family is tied to the grammar. All in all, I argued for a categorical and ultimately binary distinction between two categories of acceptability ratings that I proposed to interpret as reflecting two categories of structures, i.e., ungrammatical and grammatical ones (Chapter 3), suggesting that strong modulations in acceptability rates reflect grammatical constraints, while mild acceptability modulations reflect non-grammatical constraints tied to the functioning of our memory system.

The claim for a categorical distinction between sentences that are grammatical and sentences that are ungrammatical is in line with the assumption adopted by most linguists since at least Chomsky 1965. However, the claim that extra-grammatical factors can both influence grammatical and ungrammatical sentences represents a partial departure from the claim that extra-grammatical factors can only affect the acceptability of grammatical sentences (e.g., Sprouse 2007a, 2007b). The argument runs as follows. It starts from two assumptions: first,
ungrammatical sentences have no licit representation; second, a necessary condition for processing effects to manifest is the existence of a mental representation of the sentence. From these two assumptions, it follows that no processing effects can be observed in ungrammatical sentences, since if ungrammatical sentences lack a representation, there is nothing that can be modulated by processing factors. In support of this view, Sprouse showed that the satiation effect (i.e., the observation that acceptability rates tend to improve after repeated exposure) reported by Snyder (2000) for wh-islands and Complex NP Constraint islands was actually the result of a methodological confound, in line with the hypothesis that extra-grammatical effects such as satiation can only be found in grammatical sentences (e.g., Luka and Barsalou 2005).

However, the claim that ungrammatical sentences do not have a representation may be too strong, since there are many ungrammatical sentences that can be understood almost effortlessly. For instance, sentences containing an agreement error (e.g., The girl have left) or sentence that lack functional elements (e.g., Girl kiss boy) or sentences with an illicit word order (e.g., The student the exam pass didn’t) are ungrammatical, but there is no doubt that they can be understood, thus suggesting that the listener/reader has built a mental representation associated with these sentences. However, other syntactic violations may be harder to overcome. Freedman and Forster (1985) and Crain and Fodor (1987) put forward the idea that the understandability of ungrammatical sentences may lie in their correctability, that is on whether the sentence can be easily converted into a fully grammatical sentence. It is therefore possible that certain ungrammatical sentences cannot be represented at all, but this claim does not seem to generalize to all ungrammatical sentences. It is also possible that, when available, semantic information can facilitate, or even trigger, this repairing operation, therefore increasing the chances that the sentence can be understood. In line with this hypothesis, preliminary analyses conducted on results from Experiment 1 (Chapter 3) where animacy was manipulated in wh-islands provide indication for a satiation effect for lexically restricted wh-islands only, in line with the hypothesis that rich semantic information (such as those carried by the lexical restriction) can facilitate sentence correctability (see also Chaves and Dery 2014 for evidence that satiation effects emerge in Subject islands only when the extractee is lexically restricted). Similarly, in a speed-accuracy trade off study conducted on the four wh-island conditions investigated in Chapter 2 (i.e., Bare Identity, Inverse Inclusion, Inclusion and Complex Identity), we found that the Bare Identity condition exhibits a non-monotonical pattern, such that participants initially accepted the sentence, but their preference was reversed.
later in processing, while the other three conditions exhibit a monotonotonic pattern that gradually grows until the asymptotic level (Villata, McElree and Franck 2014). We suggested to interpret the non-monotonicity of Bare Identity as attesting for two different types of processing. On the one hand, the early acceptability raise would reflect the processing of verb subcategorization constraints. Since in our sentences the verb was always transitive (e.g., *What do you wonder who built?*), this constraint is satisfied, therefore leading participants to initially accept the sentence. On the other hand, the subsequent decrease in acceptability rates would reflect interpretative constraints: bare wh-elements (e.g., *what*) are less informative than restricted wh-elements (e.g., *which building*), which may have therefore contributed in reducing the chances to construct a meaningful mental representation associated with the sentence, thus leading participants to ultimately reject it. On this view, even if all four wh-islands conditions are ultimately ungrammatical in virtue of a principle of grammar, namely fRM, the presence of additional semantic information carried by the lexical restriction may contribute in increasing the understandability of the sentence, thus leading to an increase in acceptability rates. On the contrary, when little semantic information is available, comprehenders may fail in their attempt to construct a meaningful mental representation of the sentence, which would lead to a stronger reject.

All in all, it seems that at least certain ungrammatical sentences can be understood and, therefore, associated with a mental representation, a process that may be facilitated by the presence of additional semantic information. If so, we expect processing effects to be also observed in at least those ungrammatical sentences that can be represented, in line with what I suggested in Chapter 3, where the mild modulations observed in both ungrammatical and grammatical sentences were interpreted as stemming from processing factors. However, an alternative view that has been pursued in the literature is to interpret modulations deriving from processing factors as indicating that the source of the unacceptability of the sentence may originate from processing constraints rather than from grammatical constraints. This hypothesis has been pursued, amongst others, by Kluender and Kutas (1993), Kluender (1998), Hofmeister et al. (2007, 2010, 2013), who suggested that the origin of weak islands should be sought in processing constraints rather than in an indipendently motivated syntactic constraint. The evidence presented in this thesis does not rule out this hypothesis, even though in Chapter 3 I highlighted several challenges that are encountered by accounts trying to reduce wh-islands to processing contraints. However, the hypothesis that the source of wh-islands may
be ultimately tied in processing factors is certainly appealing under several respects. First, it is parsimonious, as it would allow to reduce a grammatical constraint to independently motivated processing factors, therefore reducing the ontology of the grammatical constraints that are supposed to be at play. Second, it would allow us to characterize all effects originating from featural similarity to a common causal source, i.e., processing. However, similar benefits would also derive from the opposite hypothesis, that is the hypotheses according to which these effects all stem from a grammatical constraint. Featural Relativized Minimality can be conceived as an attempt in this direction, since it aims at accounting for both the ungrammaticality of weak islands and the processing difficulties observed in structures such as object relative clauses with a same underlying syntactic principle, thus extending a principle of the grammar to also account for effects that have been usually conceived as processing effects. The approach I have adopted in this thesis partially departs from this proposal, since it reintroduces a distinction between two classes of effects, those stemming from the grammar and those stemming from processing, for reasons detailed in Chapter 3. It remains nonetheless desirable for future research to try to reduce this gap and reconcile effects stemming from similarity to a common causal source. The conclusions reached in Chapter 3 give rise to a clear prediction: if similarity in linguistic features such as semantic features and syntactic features not belonging to the class of criterial features is tied to the functioning of the memory system, we expect other features of the same type to modulate processing difficulties. To this aim, in Chapters 4 and 5 I focused on similarity in agreement features in fully grammatical sentences containing a long-distance dependency, such as object relatives. Similarity in terms of agreement features should not lead to ungrammaticality, but it is expected to generate increased processing difficulties (i.e., longer reading times and lower comprehension accuracy) due to increased similarity-based interference. Intriguingly, however, results from studies testing for the role of agreement features in the modulation of processing difficulties in grammatical sentences are extremely mixed, with several studies failing to show any effect of agreement features in modulating interference (e.g., Wagers et al. 2009, Dillon et al. 2013, Tanner et al. 2014, Tucker et al. 2015, Lago et al 2015), while others studies showing such an effect (in adults: Franck et al. 2015, Experiments 3 and 4 in Chapter 4 and Experiments in Chapter 5; in children: Belletti et al. 2012, Adani et al. 2010, 2014). This mixed picture is at odds with predictions from the cue-based memory model, according to which linguistic features that function as cues for the retrieval of the target (i.e., those that are represented at the integration region) should play a
role in the modulation of similarity-based interference. Results from studies failing to show an effect of similarity in agreement features are also at odds with predictions from the hypothesis that similarity can affect encoding, in which case an interference effect is expected to arise even when the overlapping features are not represented at the integration region. Finally, these results are also in contrast with predictions from Featural Relativized Minimality, according to which when agreement features belong to the phi-set of features expressed in the verbal morphology, they should play a role in modulating intervention effects (Belletti et al. 2012).

In Chapter 4, I provided a systematic investigation of the role of agreement features in modulating similarity-based interference in grammatical French object relatives (Experiments 3 and 4). In French object relatives, the verb agrees in number and gender with the fronted object, therefore providing an excellent test case to investigate the effectiveness of agreement features during structure building operations. Results showed that dissimilarity in gender and number increased sentence comprehension. These findings are in line with the cue-based retrieval memory model and fRM, since gender and number are morphologically expressed on the past participle in French, but also with the feature overwriting hypothesis.

Investigating object-past participle agreement in French was of interest for an additional reason: this agreement configuration allows us to test for the effectiveness of an object agreement retrieval cue. This is critical because of a standard assumption of the cue-based memory model, according to which the to-be-retrieved element in object relatives is the distant object, while the subject, being in the vicinity of the verb, does not need to be retrieved (e.g., McElree et al. 2003, McElree 2006). Yet, all studies investigating the effectiveness of agreement cues in the resolution of long-distance dependencies as object relatives manipulated subject agreement cues, that is cues that do not point directly to the element that is supposed to be the target of retrieval (the object). As a consequence, one may wonder whether this may have contributed in reducing the effectiveness of agreement cues at retrieval, thus explaining why several studies failed to found an effect (e.g., Wagers et al. 2009, Lago et al. 2015). However, this hypothesis appears to be invalidated by the study by Franck et al. 2015, which attested to an effect of mismatch when a subject agreement cue was manipulated (see also Experiments 2 in English, Chapter 5). In addition, results from Experiments 3 and 4 in Chapter 4 do not provide evidence for a stronger effectiveness of object retrieval cues during retrieval, since the mismatch effect in online measures was even weaker that the one found by Franck et al. (2015), thus casting doubts on the hypothesis that subject retrieval cues may be not as effective as object retrieval.
cues (even if we cannot exclude that the weaker effect found in object-past participle agreement configurations was due to the optionality of this agreement configuration in colloquial French). An additional reason that cast doubts on the assumption that only the moved object is the target of retrieval is that this would generate the wrong prediction when integrated in a cue-based retrieval model such as ACT-R. Indeed, ACT-R assumes that the target is retrieved based on its activation, which is partly determined by the strength of the association between the retrieval cues and the featural specification of the target. Under the hypothesis that the target of retrieval in object relatives is only the object, a subject retrieval cue should lead to a detrimental effect of feature mismatch, because when the object and the subject have different features, the subject retrieval cue would mismatch the feature of the target (the object), therefore leading to a decrease in its activation, thus decreasing its retrieval probabilities. However, this effect is not observed (e.g., Wagers et al. 2009, Lago et al. 2015, Franck et al. 2015, Experiment 2 in Chapter 5). A hypothesis that seems more in line with the empirical results is that both the object and the subject should be ultimately integrated with the verb, as assumed by the self-organized sentence processing model. On this view, both object retrieval cues and subject retrieval cues facilitate sentence processing in conditions of feature mismatch, as suggested by results reported in Chapters 4 and 5, together with those from the literature.

Results from the two self-paced reading experiments in Chapter 4 also reported a comparable facilitatory mismatch effect in object and subject relatives, an observation that was interpreted as indicating that the interference effects observed in both structures arise at encoding, since in subject relatives no retrieval is needed at any time. Hence, the question that arose next was what memory operations are affected by similarity. The four on-line comprehension studies presented in this thesis (Experiments 3 and 4 in Chapter 4 and Experiments 1 and 2 in Chapter 5) all converge in providing substantial support for encoding interference in long-distance dependencies, while only weak evidence for retrieval interference was observed (on-line results of Experiment 2 in Chapter 5 only, a result that was restricted to the right tail of the reading times distribution). In Chapter 5, I discussed that evidence for encoding interference also extends beyond studies manipulating agreement features. Moreover, clear evidence for retrieval interference only comes from two studies (Van Dyke and McElree 2006; Belletti et al. 2012), while several studies that have been traditionally interpreted as bringing evidence for retrieval interference are actually also compatible with encoding interference. Hence, the picture that seems to emerge is that evidence for a backward-looking cue-based retrieval mechanism in the
resolution of long-distance dependencies is not as strong as one would expect under the classical assumption that cue-based retrieval is the main memory operation that the parser has to accomplish in order to compute these structures (e.g., Lewis and Vasishth 2005; McElree 2000, 2006). In Chapter 5, I discussed how a mechanism generating encoding interference could be implemented in ACT-R. I also discussed how a self-organizing parser could account for encoding and retrieval interference effects observed in this thesis as well as in the literature. Here I want to discuss another group of theories that have challenged the very same assumption that retrieval interference is the primary cause of processing difficulties.

This hypothesis is pursued by expectation-based and production-based theories, which assume that the parser proceeds forward driven by expectations, rather than backward driven by retrieval. When the parser’s expectation is not met (i.e., when the surprisal associated with a given word is greater than zero), processing difficulty arises. According to expectation-based theories (Hale 2001; Levy 2008, 2013; Smith and Levy 2013), the parser deploys a forward predictive mechanism allowing it to anticipate upcoming material at multiple levels (syntax, morphology, semantics), such that faster processing time is associated with highly expected elements. Production-based models (e.g., Pickering and Garrod 2013, MacDonald 2013, Dell and Chang 2014) maintain that language production processes are used to anticipate upcoming linguistic material during comprehension, under the hypothesis that predicting a word is fundamentally the same as planning to say a word (Pickering and Garrod 2014). These theories offer useful insights on the mode of functioning of the parser and are able to capture a wide variety of observations. Even though these theories have been developed independently from memory-based theories and posit the accent on different components of the parser, the position that I will defend is that these theories should not be conceived as ultimately incompatible approaches. It is on the contrary very likely that both a forward-looking component and a backward-looking component are at play during sentence comprehension, therefore suggesting that, in the long run, these theories should be ultimately harmonized in a unified account of sentence processing.

One of the first evidence in support of a predictive process at play during sentence comprehension is provided by Altmann and Kamide (1999), who showed, in a visual-world paradigm, that while hearing a sentence like The boy will eat the cake, listeners directed their gaze to the cake in the visual scene before hearing the onset of the noun, thus demonstrating that they generate expectations about upcoming material. There is also substantial evidence that
expectations in sentence comprehension increase the more the appearance of the predicted element is delayed (e.g., Konieczny 2000; Vasishth 2003; Vasishth and Lewis 2006). For instance, in a study in German, Konieczny (2000) showed that when the subject and the verb are separated by a relative clause (die wunderschön war), reading times are faster at the verb (hingelegt) as compared to when the relative clause follows the verb:

(1) Er hat die Rose (die wunderschön war) hingelegt (die wunderschön war), und …
He has the rose (that was beautiful) laid down (that was beautiful) and …

‘He has laid down the rose that was beautiful, and …’

This result has been accounted for in terms of expectation: the expectation for the main verb is stronger after a relative clause than immediately after the subject because additional interveners become less and less likely, thus determining faster processing times when the verb is finally encountered. These effects are known as anti-locality and refer to the observation that processing difficulty is decreased by a greater amount of intervening material between two elements that should enter in a dependency (e.g., Konieczny 2000; Vasishth 2003; Vasishth and Lewis 2006).

Anti-locality effects were initially exclusively attested in head-final languages such as German (Konieczny 2000), Hindi (Vasishth 2003; Vasishth and Lewis 2006) and Japanese (Nakatani and Gibson 2008), where the verb typically comes after all its arguments. Because of this, these effects were interpreted as being the results of different processing strategies for head-final languages as compared to head-initial languages. In particular, Levy et al. (2013) suggested that speakers of verb-final languages may be more able to retain predictions of upcoming elements than speakers of head-initial languages. Consistently with this hypothesis, Vasishth et al. (2010) showed that German speakers are insensitive to the illusion of the missing-VP in multiple center embeddings, contrary to English speakers. That is, while omitting the middle verb in a double center embedding leads to processing facilitation in English, which attests to an illusion of grammaticality, it leads to increased processing difficulties in German (e.g., The carpenter who the craftsman that the peasant carried (hurt) supervised the apprentice). The authors suggested that this asymmetry attested to different parsing strategies in verb-initial and verb-final languages.

Intriguingly, however, anti-locality effects have also been recently attested in a study conducted in English by Jaeger, Fedorenko and Gibson (2005). The authors showed that when the subject and the verb are separated by several prepositional phrases, as in The player that the coach met
[near the gym] [by the river] [in Paris] bought the house (example from Jaeger et al. 2005), faster reading times at the main verb region are observed in the three prepositional phrases condition. Since English is a head-initial language, this result casts doubts on the hypothesis that anti-locality effects are exclusively due to processing strategies specific to head-final languages. Moreover, evidence for both locality and anti-locality effects within a same language as also been observed (e.g., Levy et al. 2013 for Russian and German; Husain et al. 2014 for Hindi). Taken together, these findings raise the question of knowing which are the precise conditions that lead to the manifestation of locality and anti-locality effects in sentence processing.

A promising way to address this question is provided by the model of sentence processing developed by Lewis and Vasishth (2005) and implemented in ACT-R (Anderson et al. 2004): activation-decay and interference during retrieval accounts for locality effects, while repeated activation of the to-be-retrieved element accounts for anti-locality effects. I will give an example of reactivation, since in this thesis I have only been concerned with examples of interference. Consider again the example from Konieczny (2000) reported in (1). The ACT-R model accounts for this result in terms of reactivation: the relative clause reactivates the object, which results in an activation boost of this element, thus facilitating its retrieval when the verb is encountered (Vasishth and Lewis 2006). Hence, by assuming that intermediate materials can reactivate the target, thus increasing its activation, the ACT-R model can also account for anti-locality effects.

Additional evidence for a facilitatory effect due to reactivation is provided by Dillon et al. (2013) in reflexive dependencies. These dependencies are resistant to the interference of an intervening mismatching element (e.g., Sturt 2003, Dillon et al. 2013), an observation that contrasts with the recurring finding that agreement dependencies are extremely sensitive to interference (e.g., Wagers et al. 2009, Franck et al. 2015, as well as results from Chapters 4 and 5). To account for this asymmetry, Dillon et al. suggested that in reflexive dependencies the antecedent of the reflexive is reactivated at the main verb immediately preceding the reflexive for subject-verb intergration (e.g., The new executive who oversaw the middle managers apparently doubted himself on most major decisions), which gives it an activation advantage over the intervener (the middle managers), therefore preventing similarity-based interference effects to show up.
However, despite the ACT-R model is able to account for several empirical observations, and in particular for both locality and anti-locality effects, the anti-locality effect reported from Jaeger et al. (2005) in English cannot be accounted for in terms of reactivation. In this case, the interpolated prepositional phrases do not re-activate the target element (i.e., the subject), but a facilitatory effect is nonetheless observed when the verb is reached. This finding thus challenges an interpretation in terms of reactivation, but it also challenges the hypothesis that speakers of head-initial languages struggle in maintaining predictions in memory, therefore preventing anti-locality effects to show up in these languages (Vasishth et al. 2010, Levy et al. 2013). In addition, this finding is also in contrast with results from Grodner and Gibson (2005), where longer reading times at the verb were observed when a prepositional phrase was added between a subject and its verb, thus attesting to locality effects (e.g., *The nurse (from the clinic) supervised the administrator*). However, the study from Jaeger et al. and the study from Grodner and Gibson are not directly comparable: first, Grodner and Gibson tested the interpolation of only one prepositional phrase, while Jaeger et al. tested up to three propositional phrases; second, Grodner and Gibson compared a condition in which no intervening material was present with a condition in which one prepositional phrase was added, while Jaeger et al. only tested conditions with interpolated prepositional phrases; third, while Grodner and Gibson tested the effect of an interpolated prepositional phrase in subject-verb dependencies, Jaeger et al. tested it in object relative clauses. Hence, before firm conclusions can be drawn, it is necessary to conduct studies that provide a more minimal comparison between the two experiments by testing the same material as in Grodner and Gibson and manipulating the number of intervening prepositional phrases (none, one, two or three prepositional phrases). Moreover, since in the study from Jaeger et al. the interpolated material was added between the relative verb and the matrix verb, we cannot exclude that the observed facilitatory effect may be the result of an end-of-sentence wrap up effect rather than constituting evidence for anti-locality: in the three prepositional phrases condition, the parser has extra time to deal with the resolution of the relative clause before encountering the matrix verb, which may therefore have lead to a facilitatory effect. One way to disentangle sentence wrap-up effects and expectation-based effects would be to manipulate the amount of the interpolated material both inside the relative clause (2) and at the end of it (3):

(2) The player that the coach [of the team] [from Paris] [in France] met suddenly bought the house but his wife didn't like it very much.
(3) The player that the coach met [of the team] [from Paris] [in France] suddenly bought the house, but his wife didn't like it very much.

If Jaeger et al.‘s results are the manifestation of anti-locality effects, then faster reading times should be observed in the three prepositional phrases condition both in (2) at the relative verb \textit{met}, and in (3) at the matrix verb \textit{bought}. If, on the contrary, Jaeger et al.‘s results are the result of sentence wrap-up effects, then the matrix verb should be read faster in (3) in the three prepositional phrases condition, but this effect should not be observed at the relative verb in (2), where we would expect a locality effect to manifest, with longer reading times the more additional material is added. All in all, it seems desirable for future research to cast light on how expectations and memory interact across and within languages by investigating which conditions lead memory-based effects to prevail over expectation-based effects and \textit{vice versa}, thus contributing in advancing our understanding of language comprehension.

Another important aspect that deserves further investigation concerns the role of encoding in sentence processing. In Chapter 5, I discussed that current models of sentence processing must include a mechanism that generates encoding interference effects, and I discussed a possible implementation in both ACT-R and SOSP. However, encoding effects seem difficult to be captured through the expectation-based theories discussed in this section. Although it seems plausible to assume that what has been encoded will shape our predictions about what will come next, it is unclear how a mismatch in the featural specification of two noun phrases may facilitate expectations. First, in structures such as object relatives, if expectations were at play in predicting verb’s agreement based on the agreement features of the arguments of the verb, then we would expect the opposite effect to show up, that is a facilitatory effect in match conditions, similarly to what is observed in production (Chapter 4). Indeed, when both noun phrases have the same agreement feature, even if the wrong agreement controller is selected, the agreement feature on the verb will never dash the expectation of the parser. Second, encoding effects are observed even when the verb does not agree with the controller, in which case it is rather unclear how expectations-based theories could account for this result.

In Chapter 5, I have suggested that a self-organized sentence model (SOSP; e.g., Tabor and Hutchins, 2004) can account for both encoding and retrieval effects with a same mechanism. However, in this last chapter I also discussed evidence for anti-locality effects, which enlarges the realm of observations that need to be accounted for by a theory of sentence processing. The question arises as to whether SOSP can also account for anti-locality effects.
One of the defining properties of SOSP is that the strength of the links between treelets grows as a function of the feature match between the treelets, the passage of time and additional consistent linkages. A consequence of this architecture is that the stronger a link has become, the harder it is to undo, an effect known as digging-in (Tabor and Hutchins 2004; see also Ferreira and Henderson 1991). Tabor and Hutchins tested sentences such as As the author wrote (the essay) the book grew and As the author wrote (the essay) the book describing Babylon grew, where both the transitivity of the sentence and the length of the ambiguous region were manipulated. They observed lower comprehension accuracy and longer reading times at the verb grew in the intransitive long condition as compared to the intransitive short condition. They interpreted this result as follows: the erroneous link between the verb wrote and the noun phrase the book in the intransitive condition becomes stronger when the ambiguous region is long than when it is short, thus becoming harder to undo. Evidence in support of digging-in effects has also been found in a recent study that I have conducted in Italian relative clauses (Villata, Franco and Lorusso 2017). Italian relative clauses are ambiguous between a subject and object reading since Italian permits post-verbal subjects. Hence, in Italian, a sentence like The girl that draws the clown smiles (La bambina che disegna il pagliaccio sorride) is fully ambiguous between a subject relative clause (The girl that is drawing the clown is smiling) and an object relative clause (The girl that the clown is drawing is smiling). In the absence of cues that signal that the sentence must be interpreted as an object relative, the parser engages in a subject analysis. Two grammatical cues have been shown to efficiently trigger an object reanalysis: i) word order (OSV), and ii) number agreement (e.g., OSG VPL SPL) (e.g., Arosio et al. 2009; Guasti et al. 2012). However, both the word order cue and the number agreement cue arrive at early processing stages, namely at the relative clause. In a self-paced reading study accompanied with a sentence comprehension task targeting thematic roles attribution, we tested the effectiveness of a gender cue appearing late in processing, namely after the relative clause, in triggering reanalysis. We tested sentences like Il sindaco-M che consulta la giornalista-F prima di essere ascoltat-o-M/a-F da tutti vive a Parigi (The mayor-M that consults the journalist-F before being heard-M/F by everyone lives in Paris): when the past participle (heard) agrees in gender with the second noun phrase (the journalist), the sentence must be interpreted as an object relative clause, thus leading the parser to revise its initial interpretation. Results showed that participants were both slower at the region following the past participle and less accurate in answering the comprehension question when the past participle agreed with...
the second noun phrase, thus requiring an object analysis, than when it agreed with the first noun phrase, thus supporting the subject analysis. In particular, comprehension accuracy showed that participants failed to access the object analysis 80% of the time. However, the percentage of correct responses in the object condition did not differ from the percentage of incorrect responses in the subject condition, thus providing no evidence that participants ever truly accessed the object analysis. Hence, while previous studies manipulating early disambiguating cues showed that comprehenders successfully reanalyse the sentence (e.g., Arosio et al. 2009; Guasti et al. 2012), we provide evidence that when the cue arrives later in processing, no reanalysis is triggered, in line with SOSP.

The digging-in effect is naturally captured by SOSP: since the linkages progressively become stronger, to discard an expectation that has been maintained longer is costlier than to discard an expectation that has been maintained for a shorter period of time. This is because in SOSP when the parser builds up the structure, the already activated treelets “project into the future”, in the sense that they predict future treelets to combine with. When a treelet is expected, it starts becoming active. For instance, when a noun is encountered, not only a noun phrase treelet will be activated, but also a verb phrase treelet will become activated, due to the fact that a noun needs a verb to combine with. As more material arrives, this expectation will be strengthened, and so the activation of the expected treelet.

At this point, it should be clear how SOSP could account for anti-locality effects. If when the input is incompatible with an expectation that has been maintained for long time the system incurs in a processing cost (digging-in effect), then the reverse holds true as well: when the input is compatible with an expectation that has been maintained for long time, this will result in a processing facilitation (anti-locality). Hence, anti-locality effects would be the result of the high activation level reached by the expected element as a consequence of the parser’s predictions. When the expected element is finally encountered, the linkage between its treelet and the structure that has already been built will happen very fast. This will manifest in faster reading times, under the assumption that reading times reflect the linkage speed. If this proposal is somehow on the right track, it may pave the way for a yet another possible unification of both locality effects and anti-locality effects under a same theory of sentence processing, a perspective that would be worth exploring in future research.

Of course, many questions remain unanswered. For instance, it remains to determine how and when the retrieval and the predictive components of the parser come into play in sentence
comprehension, under which conditions one is likely to prevail over the other and if these patterns are language-specific or rather reflect universal constraints.

Also, it appears important to delve more into the question of whether anti-locality effects are observable also in head-initial languages, possibly along the line suggested in this chapter, and how we can reconcile anti-locality effects with syntactic principles of locality. Moreover, should we assume two independent syntactic principles to account for, respectively, locality and anti-locality effects, or can these effects be captured with a single grammatical principle? It also remains an open question whether featural similarity effects can be reduced to a unique underlying principle, and if this principle would be ultimately a principle of grammar or processing.

Further research is also needed to clarify whether encoding and retrieval interference arise from fundamentally distinct processing mechanisms or if they are generated by a single mechanism, in which case the distinction between encoding and retrieval interference would vanished.

Also, the recurrent findings presented in this thesis showing a stronger effect of feature mismatch in off-line than in on-line measures also deserve further investigation: what are the processes at play during sentence comprehension answering? Are these processes similar in kind to those that are supposed to be at play in the real-time parse of the sentence or do they tap into ultimately different processes?

This dissertation, I hope, represents a contribution in our understanding of syntactic complexity and the mechanisms underlying sentence processing, paving the way for future research.
References


survey tool /LimeSurvey Project Hamburg, Germany. URL http://www.limesurvey.org


