EXPLAINING THE NULL SUBJECT IN THE ROMANCE LANGUAGES:
FINDING A SUITABLE FRAMEWORK

BY

PETER S. MANASANTIVONGS
PETERHOUSE
UNIVERSITY OF CAMBRIDGE

APRIL 2004

THIS DISSERTATION IS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
DECLARATIONS

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text.

This dissertation does not exceed the word limit as permitted by the Degree Committee of the Faculty of Modern and Medieval Languages.

This dissertation submitted is not substantially the same as any that I have submitted for a degree or diploma or other qualification here or at any other university.

Peter S. Manasantivongs
April 2004
ABSTRACT

This dissertation challenges the commonly held belief that subject pronouns are always optional in the Romance languages. Through empirical data collected on Brazilian Portuguese, European Peninsular Spanish, Catalan, European Portuguese, Italian, and Puerto Rican Spanish, it investigates the effects of gender, number, and person as distinguishing features between antecedents; the topichood of the antecedent; the distance between a subject and its antecedent; and gender-inherent predicates, on the obligatoriness of overt subject pronouns.

There are four primary goals of this work. First, it surveys a selected sample of treatments of the pro-drop phenomenon in various grammars and syntactic textbooks to examine the extent to which the claim that subject pronouns are never required has pervaded discussions of Romance null subject usage. Second, it attempts to show how several phenomena that could be expected to affect pro-drop, such as grammatical relations, thematic roles, the animacy hierarchy, and distinguishing verbal desinences, do not in fact have any direct relevance. Third, it offers a systematic account, modelled primarily on work done by Samek-Lodovici (1996) and Cole (2000), of when overt subject pronouns are required in the various linguistic environments of the six Romance varieties under consideration. Fourth, it proposes, based on the data provided, that Optimality Theory (OT) is the best approach in which to describe Romance null subjecthood.

It is argued that the traditional Principles and Parameters Theory (PPT) framework, in both its Government and Binding (GB) and Minimalist Program (MP) formats, is largely inadequate in explaining the microvariation witnessed in Romance subject pronoun requirements, both language-Internally and cross-linguistically. Whereas the customary approach in PPT is to label languages as either [+pro-drop] or [-pro-drop], it is asserted that a more accurate representation of the null subject phenomenon is achieved in OT by postulating that the various interactions of the competing core constraints DROP and PARSE are best able to reflect the observed differences in overt subject pronoun usage. This is particularly evident intra-linguistically, but is equally true of inter-linguistic comparisons as well.
ACKNOWLEDGEMENTS

My most heartfelt gratitude goes to Dr Adam Ledgeway, my supervisor, for his encouragement, support, and faith in me over the past three years. Adam took a chance on a neophyte linguist in 2001 and agreed to supervise this doctoral dissertation. He never hesitated to give me as much time as I needed, whenever I needed it, to discuss questions, issues, problems, and obstacles. His contagious enthusiasm, which he dispensed liberally, was exactly the right boost, both to motivate me during those frustrating stretches, and also to inspire me even further when things were going smoothly. Adam has always shown a genuine personal interest in me and my work, for which I am truly thankful. I could not have asked for a better supervisor, and am fully aware of just how fortunate I am.

I would like to express my sincere appreciation to the Gates Cambridge Trust; the Department for Education and Skills (DfES) and Universities UK for the Overseas Research Students Awards Scheme (ORS); the Cambridge Overseas Trust; and the John Finley Travelling Fellowship at Eliot House, Harvard University; for their financial support. They have made my studies here possible, providing me the invaluable opportunity to study full-time and in greater depth a that I had previously only been able to explore as a dabbler. This is a gift that I shall not soon forget.

Without my data, there is no dissertation. My informants were generous with their time and patience throughout my repeated requests for grammatical judgements and probing clarifications. I am grateful to all those who provided me with the information that has formed the foundation of this work, but especially: Luisa Astruc-Aguilera, Anna Bayona Font, Vitor Bernardes Pinheiro, Patricia Boulhosa, Luis Briseño-Roa, Rita Freire Faria, Chiara Frenquellucci, Casilda García de la Maza, Ana Lourenço, Biagio Mazzi, Marisa Perez, Antonio Vale, and Cristiana Viegas-de-Andrade.

Finally, earnest thanks go to my family and friends for their love and camaraderie. My experiences in Cambridge, both academic and personal, would not have been this fulfilling without them. I wish I had the space to list everyone by name, but I would like to mention in particular Drs Alan Abrahams, Theresa Biberauer, Chiara Frenquellucci, Jennifer Shea, and Dalal Tahri, for their persistent efforts in convincing me that there was indeed a dissertation somewhere inside me, and for telling me repeatedly to keep at it, and not to quit. As tempted as I was at times to abandon this project, I am glad that I decided to heed their wise words instead and take them to heart. Part of my joy in completing this assignment comes from having proved them correct.
# Table of Contents

Chapter 1: Framing the problem  

1.1 The state of discussions on *pro*-drop  
1.2 Understanding the phenomenon  
1.2.1 What is *pro*-drop?  
1.2.2 Is it *pro*-drop?  
1.2.3 How does a language “stop” being *pro*-drop?  
1.3 What the grammars say, and do not say  
1.3.1 Latin  
1.3.2 Italian  
1.3.3 Portuguese  
1.3.4 Spanish  
1.3.5 French  
1.3.6 Catalan  
1.3.7 Galician  
1.3.8 Occitan  
1.3.9 Sardinian  
1.4 Goals of this work  
1.4.1 Are subject pronouns always optional?  
1.4.2 Is *pro*-drop uniform within Romance?  
1.4.3 A null subject “parameter”?  

Chapter 2: Identifying what does not affect the licensing of *pro*  

2.1 Investigating what is not relevant to this study  
2.2 Grammatical relations and thematic roles  
2.2.1 Canonical relationships  
2.2.2 Switch reference  
2.2.3 The grammatical relations hierarchy  
2.2.4 Testing the factors  
2.3 Animacy  
2.3.1 The animacy hierarchy  
2.3.2 The role of animacy in syntax  
2.3.3 Salience and topicality  
2.3.4 Testing the factors  
2.4 Subject pronoun usage in the history of the French language  
2.4.1 Word order and null subjects in Old and Middle French
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4.2</td>
<td>Switching of a parameter</td>
<td>55</td>
</tr>
<tr>
<td>2.4.3</td>
<td>Verbal morphology and assumptions</td>
<td>57</td>
</tr>
<tr>
<td>2.4.4</td>
<td>Modern Carribean Spanish: déjà vu?</td>
<td>60</td>
</tr>
<tr>
<td>2.5</td>
<td>Summary</td>
<td>62</td>
</tr>
<tr>
<td>Chapter 3</td>
<td>Literature review and methodology</td>
<td>64</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>64</td>
</tr>
<tr>
<td>3.2</td>
<td>Gilligan (1987)</td>
<td>65</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Null subjects, subject inversion, and <em>that</em>-trace filter violations</td>
<td>65</td>
</tr>
<tr>
<td>3.2.2</td>
<td>The correlation between null subjects and rich agreement</td>
<td>67</td>
</tr>
<tr>
<td>3.2.3</td>
<td>A <em>pro</em>-drop typology</td>
<td>69</td>
</tr>
<tr>
<td>3.2.4</td>
<td>The role of parameters in describing the null subject</td>
<td>71</td>
</tr>
<tr>
<td>3.2.5</td>
<td>The influence of features</td>
<td>72</td>
</tr>
<tr>
<td>3.3</td>
<td>Sigurðsson (1993)</td>
<td>73</td>
</tr>
<tr>
<td>3.4</td>
<td>Samek-Lodovici (1996)</td>
<td>75</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Using Optimality Theory to understand syntax</td>
<td>75</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Questioning <em>pro</em></td>
<td>78</td>
</tr>
<tr>
<td>3.4.3</td>
<td>Constraints proposed</td>
<td>81</td>
</tr>
<tr>
<td>3.4.4</td>
<td>Constraint interactions, their rankings, and cross-linguistic variation</td>
<td>84</td>
</tr>
<tr>
<td>3.4.5</td>
<td>Points of further research</td>
<td>88</td>
</tr>
<tr>
<td>3.5</td>
<td>Speas (1997)</td>
<td>90</td>
</tr>
<tr>
<td>3.6</td>
<td>Baković (1997)</td>
<td>95</td>
</tr>
<tr>
<td>3.7</td>
<td>Cole (2000)</td>
<td>100</td>
</tr>
<tr>
<td>3.7.1</td>
<td>Default interpretations</td>
<td>101</td>
</tr>
<tr>
<td>3.7.2</td>
<td>Morphological maximality</td>
<td>104</td>
</tr>
<tr>
<td>3.7.3</td>
<td>Semantic identification</td>
<td>107</td>
</tr>
<tr>
<td>3.7.4</td>
<td>Ariel (1988; 1990) and Accessibility Theory</td>
<td>109</td>
</tr>
<tr>
<td>3.7.5</td>
<td>Cole’s comments on Samek-Lodovici (1996)</td>
<td>113</td>
</tr>
<tr>
<td>3.7.6</td>
<td>Cole’s question types</td>
<td>117</td>
</tr>
<tr>
<td>3.8</td>
<td>The present research</td>
<td>122</td>
</tr>
<tr>
<td>3.8.1</td>
<td>How it differs from previous studies</td>
<td>123</td>
</tr>
<tr>
<td>3.8.2</td>
<td>Methodology</td>
<td>125</td>
</tr>
<tr>
<td>Chapter 4</td>
<td>Presentation of the data and initial observations</td>
<td>129</td>
</tr>
<tr>
<td>4.1</td>
<td>Topic followed by Main Clause</td>
<td>130</td>
</tr>
<tr>
<td>4.1.1</td>
<td>No features differentiating the two competing antecedents</td>
<td>130</td>
</tr>
</tbody>
</table>
4.1.2 Gender differentiating the two competing antecedents 131
4.1.3 Number differentiating the two competing antecedents 133
4.1.4 Person differentiating the two competing antecedents 134
4.1.5 Multiple features differentiating the two competing antecedents 135

4.2 Topic followed by Main Clause and Subordinate Clause 138
4.2.1 No features differentiating the two competing antecedents 138
4.2.2 Gender differentiating the two competing antecedents 139
4.2.3 Number differentiating the two competing antecedents 140
4.2.4 Person differentiating the two competing antecedents 141

4.3 Two sentences 143
4.3.1 No features differentiating the two competing antecedents 144
4.3.2 Gender differentiating the two competing antecedents 145
4.3.3 Number differentiating the two competing antecedents 146
4.3.4 Person differentiating the two competing antecedents 147

4.4 Gender-inherent predicates 149

4.5 Exploratory generalisations 151

Chapter 5: The validity of PPT approaches 156

5.1 Limitations of the Government and Binding framework 156
5.1.1 Seeking out this alleged “pro-drop parameter” 156
5.1.1.1 Chomsky’s definitions 156
5.1.1.2 Jaeggli and Safir’s (1989a) “morphological uniformity” 158
5.1.1.3 Rizzi (1986) 161
5.1.1.4 Contemporary descriptions of null subject usage 162
5.1.2 Attempting to fit the data into a GB model 165
5.1.2.1 Setting up a hypothetical framework 165
5.1.2.2 The proposed model at work 167
5.1.2.3 How to analyse and compare in this model 170
5.1.2.4 The possibility of feasible substitute models 172
5.1.2.5 Using parameters to explain a non-parametric phenomenon? 174

5.2 Limitations of the Minimalist framework 175
5.2.1 EPP and feature strength 175
5.2.2 Attempting to fit the data into a MP model 179
5.2.3 Looking into the feature bundles 182

Chapter 6: Appealing to Optimality Theory 190

6.1 Advantages over Principles and Parameters 190
6.1.1 Getting rid of parameters 190
6.1.2 Making clear our stance
6.2 Pre-existing null subject explanations in OT
   6.2.1 Deficiencies in Samek-Lodovici’s (1996) model
   6.2.2 Deficiencies in Cole’s (2000) model
6.3 A new proposal
   6.3.1 Defining the constraints
      6.3.1.1 DROP(GENDER), DROP(NUMBER), DROP(PERSON), and DROP(NO_DIFF)
      6.3.1.2 PARSE(PRO), PARSE(C), and PARSE(S)
      6.3.1.3 A disclaimer on grammaticality, optionality, and Optimality Theory
6.3.2 Introduction of constraints
6.3.3 Using the constraints to understand the data
   6.3.3.1 Brazilian Portuguese
   6.3.3.2 Castilian
   6.3.3.3 Catalan
   6.3.3.4 European Portuguese
   6.3.3.5 Italian
   6.3.3.6 Puerto Rican Spanish
6.3.4 General comments about the model
   6.3.4.1 Presence of multiple features
   6.3.4.2 Treating co-reference with a topic antecedent in our model
   6.3.4.3 Ranking the languages and features
6.4 Unresolved issues
   6.4.1 Differences in the strengths of features
   6.4.2 Cole’s “semantically adequate identification”
   6.4.3 Cole’s “morphologically maximality” and licensing mechanism of pro
   6.4.4 Ariel’s “accessible antecedent”
   6.4.5 Gender-inherent predicates: a second look
   6.4.6 Critiques of OT syntax

Chapter 7: Conclusion

7.1 Summarising the present work
7.2 Differences, groupings, and constraint hierarchies: some new insights
   7.2.1 Determining the identity of each language
   7.2.2 Pinpointing differences
   7.2.3 Thinking about traditional alignments
7.3 Areas of future research
   7.3.1 More languages
   7.3.2 More linguistic environments

Appendices
References
Chapter 1: Framing the problem

1.1 The state of discussions on pro-drop

Characterisations of the pro-drop phenomenon are frequently inadequate because they are often too simplistic and general; as a result, they usually fail to demonstrate how complex the situation actually is. Instances of this overwhelmingly cursory treatment of the null subject are easily found in linguistic textbooks and references:

(1) "[Pro-drop] [I]languages do not require that the subjects of finite clauses be overt." (Encyclopedia of Languages and Linguistics, Atkinson (1993: 2855))

(2) "Pronominal subjects are optional in Italian finite clauses. They can be present or absent." (Introductory level syntax textbook, Ouhalla (1999: 311))

This descriptive simplification is not restricted to works covering a wide range of topics or to those that address the audience at an elementary level. We find equally broad descriptions in grammars and in works of a technical and specific nature, as the following quotes show:

(3) "Catalan is characterized, like Spanish, Italian, and Portuguese, but unlike French, by the way in which subject pronouns accompany verbs only for particular emphasis." (Catalan grammar, Wheeler, Yates, and Dols (1999: 160))

(4) "Spanish, like a number of other Romance languages, for example, Italian, Portuguese, does not require the presence of a subject pronoun [...]" (Stewart (1999: 108)).

Upon quick consideration, the four quotes above seem to be uncontroversial. However, after some deeper scrutiny, we see that each is either slightly flawed or makes a claim that is ambiguous or presumptuous. For example, despite the assertion in (1), we shall see in greater detail in Section 1.3.2 that the pronoun tu must generally...
be phonetically expressed in Italian subordinate clauses with a subjunctive verb, if a second person singular subject is intended. Ungrammaticality arises from the omission of the explicit subject pronoun and, therefore, it is not always the case that overt subjects in finite clauses are not required. Furthermore, it would be imprudent to assert that Italian is not a pro-drop language. Clearly, then, there is some sort of discrepancy in statements such as (1) that needs to be resolved.

It is worthwhile to examine what exactly is intended by the word “optional” in (2). While the previous paragraph has already hinted that pronominal subjects must be present in certain linguistic environments, one should question whether the use of subject pronouns is truly optional in the sense that their presence or absence does not make any difference to the communication and the message of the utterance. If overt pronouns are indeed optional, and no difference results whether they are realised or not, then why should speakers ever bother using them, if they add absolutely no meaning to the utterance? Conversely, if subject pronouns are used only for emphasis, as (3) asserts, then their appearance in a sentence serves a genuine linguistic function, just as, for instance, tones do in certain languages. If emphasis cannot be attained without the use of an overt subject in some Romance languages, just as the distinction between otherwise homophonous words in, say, Mandarin, cannot be discriminated without the use of tone, then it would be illogical to argue that subject pronouns are optional. No one would argue that tone is “optional” in Mandarin. The assumption that usage of an overt subject pronoun is a choice without consequence is a simplification that has appeared frequently in descriptions of pro-drop languages.

(3) and (4) exemplify the widely-held conception that the (non-)usage of subject pronouns in all the different Romance languages can be summarised by one, all-encompassing umbrella known as “pro-drop”. While it is perhaps convenient to label all these languages as “pro-drop” and others such as French and the Germanic languages as not being “pro-drop”, these sorts of blanket statements fail to acknowledge the different grammatical rules in effect in the various Romance varieties, and hence make the presumption that one phenomenon, this “pro-drop”, can adequately explain what we observe in Spanish, Italian, and other languages.

---

1 While this is generally the case, it should be pointed out that some Germanic varieties do allow overt pronouns. We discuss this more in Chapter 3 with reference to Bavarian German.
Maybe this is true; perhaps it indeed is the case that one set of pro-drop rules is able to account for all the restrictions on the appearances and absences of subject pronouns in the Romance languages. One of the goals of the present research is to figure out whether this master key exists. However, as of yet, there has been no sufficient explanation of why, for example, overt subjects seem to be used quite frequently in certain Latin American varieties of Spanish and Portuguese, but apparently not nearly as often in their respective European varieties; or, even within one language, why, for instance, only Italian second person singular subjunctive verbs require an overt subject. Until the linguistic community has succeeded in identifying how the various instances of pro-drop are related among the Romance languages, it may be considered a bit overeager at the present stage to clump Spanish, Portuguese, and Italian together, as is often done, or to state that one language acts just like its sister languages, as (3) and (4) above show.

Not only is it slightly misleading at this point to assert that all these Romance varieties function in the same manner with respect to null subjects, but also claims such as those in (3) and (4) fail to take into account the numerous variations that exist throughout the world even within one language. Spanish, for example, is manifested in myriad varieties throughout Latin America and even within Spain itself, and while it is perhaps convenient to squeeze all of these versions neatly under one heading of [+pro-drop], thereby making a blanket statement about the entire group, it may be erroneous to suggest that, as we shall explore in greater depth later in this work, particular varieties of Spanish, notably those spoken in the Caribbean, allow and may even require the use of subject pronouns. It is very common in some of these Spanish varieties to express the subject pronoun, so much so as to violate certain prescriptive rules of its usage. It may very well be the case then that these overt subject pronouns are obligatory in the accepted, colloquial usage of these varieties.

The sampling of quotes at the beginning of this section serves to highlight an important point in the discussion on pro-drop. There is a tendency to oversimplify descriptions of the pro-drop phenomenon by ignoring the fact that the situation is not as simple and clear-cut as it is often made out to be. As a result, the intricacies of precisely when pronominal subjects can be dropped, and when they cannot be omitted, are not fully appreciated. This has non-trivial ramifications. For instance, it may require us to re-examine how to characterise these overt subjects in certain contexts; perhaps they are more accurately recognised as subject clitics rather than as
subject pronouns. If this is the case, and we accept Roberge’s proposition that “languages with subject clitics are never pro-drop in the traditional sense of the word” (1986: 55), then it forces us to question what pro-drop is exactly\(^2\). Consequently, the data might lend substantial support to the suggestion that these subject pronouns be treated as part of the verbal morphology, and thus part of the verbal paradigm. From a more tangible perspective, this information would be valuable in higher-level foreign language instructional textbooks, and for those who have an extremely strong grasp of the language but who seek to advance their knowledge of it towards that of a native speaker. So it is with both theoretical and practical goals in mind that there is significant scope to pursue a thorough account of the pro-drop phenomenon in the Romance languages.

1.2 Understanding the phenomenon

1.2.1 What is pro-drop?

There have been several approaches towards explaining this concept over the previous thirty years, each with its merits and shortcomings. What follows here is a brief synopsis of each of the frameworks, along with their ramifications on the current state of analyses on pro-drop, as well as the weaknesses they display that prompt us to consider a more complete and accurate description of the null subject phenomenon.

Taraldsen (1978) was an early proponent of the theory that licensing of null subjects could be attributed to the richness of verbal morphology. Of course, subsequent studies have shown that counterexamples exist in both directions. For instance, German and Icelandic verbs inflect for person, number, and tense, but null referential subjects are not permitted in these languages\(^3\), whereas in Chinese and Japanese, where verbs do not inflect at all for number or person, both referential and expletive subjects are allowed to be covert. Clearly, then, the correlation between

\(^2\) But see Rizzi (1986) for his argument that there is no correlation between subject clitics and a language’s null subject parameter setting.

\(^3\) It is believed that this is due to these languages’ status as V2 languages. While data and observations from V2 phenomena will be discussed in this work where relevant, it is beyond the scope of the present research to investigate in depth the relation between V2 languages, inflection, and null subjects. See Vikner (1995) for detailed coverage of this issue.
subject pronouns. If this is the case, and we accept Roberge’s proposition that “languages with subject clitics are never pro-drop in the traditional sense of the word” (1986: 55), then it forces us to question what pro-drop is exactly\(^2\). Consequently, the data might lend substantial support to the suggestion that these subject pronouns be treated as part of the verbal morphology, and thus part of the verbal paradigm. From a more tangible perspective, this information would be valuable in higher-level foreign language instructional textbooks, and for those who have an extremely strong grasp of the language but who seek to advance their knowledge of it towards that of a native speaker. So it is with both theoretical and practical goals in mind that there is significant scope to pursue a thorough account of the pro-drop phenomenon in the Romance languages.

### 1.2 Understanding the phenomenon

#### 1.2.1 What is pro-drop?

There have been several approaches towards explaining this concept over the previous thirty years, each with its merits and shortcomings. What follows here is a brief synopsis of each of the frameworks, along with their ramifications on the current state of analyses on pro-drop, as well as the weaknesses they display that prompt us to consider a more complete and accurate description of the null subject phenomenon.

Taraldsen (1978) was an early proponent of the theory that licensing of null subjects could be attributed to the richness of verbal morphology. Of course, subsequent studies have shown that counterexamples exist in both directions. For instance, German and Icelandic verbs inflect for person, number, and tense, but null referential subjects are not permitted in these languages\(^3\), whereas in Chinese and Japanese, where verbs do not inflect at all for number or person, both referential and expletive subjects are allowed to be covert. Clearly, then, the correlation between

\(^2\) But see Rizzi (1986) for his argument that there is no correlation between subject clitics and a language’s null subject parameter setting.

\(^3\) It is believed that this is due to these languages’ status as V2 languages. While data and observations from V2 phenomena will be discussed in this work where relevant, it is beyond the scope of the present research to investigate in depth the relation between V2 languages, inflection, and null subjects. See Vikner (1995) for detailed coverage of this issue.
recoverability of subjects in the verbal morphology and their ability to be phonetically absent is by no means direct.

Chomsky (1981: 240) identified a cluster of properties that, he claimed, is normally associated with pro-drop languages. They are listed below, with corresponding examples from Portuguese:

(i) missing subject

\[ \text{Parti às sete.} \]
‘I left at 7 o’clock.’

(ii) free inversion in simple sentences

\[ \text{Fê-lo o Henrique.} \]
‘Henry did it.’

(iii) long wh-movement with apparent violations of the \*[that-t] filter

\[ \text{Quem é que achas que chegou ontem?} \]
‘Who do you think it is that arrived yesterday?’

(iv) empty resumptive pronouns in embedded clauses

\[ \text{A Maria disse que [ ] ia protestar contra à situação.} \]
‘Maria said that she was going to protest the situation.’

While he maintained at the time that “when this [pro-drop] parameter is set one way or another, the clustering of properties should follow” (Chomsky 1981: 241), it is generally believed that this is not a completely accurate representation of the pro-drop phenomenon.

Take for instance (ii), the possibility of free inversion in simple sentences. Safir (1982) (cited in Roberge (1986: 58)) challenges this point by demonstrating that Trentino and Modenese, two Italian dialects, are not missing-subject languages while
having free inversion, whereas Portuguese permits null subjects but forbids free inversion.

A fifth property alongside the four listed above is invoked by some as well (e.g., Kempchinsky (1984:4)):

(v) empty expletive elements (cf. Chomsky (1981: 281); Rizzi (1986: 410))

[ Parece que vai chover hoje.
‘It appears that it is going to rain today.’

The problem with this addition, of course, is that it renders the entire cluster of properties unaccountable for languages with different parameter settings for null expletive and for null referential subjects, as we shall see below with Icelandic and Neapolitan. So it is clear that an investigation is needed to modify this cluster of properties; we will attempt to clarify which properties really do correspond to null subject languages, and which are inconsistently linked to them. This does not preclude, of course, the addition of new properties to this cluster if evidence is presented to support their inclusion.

A significant breakthrough in the typology of null subject languages was attained when Rizzi (1982: 143) identified classes of null subject languages based on their settings of two distinct parameters: the possibility of having null expletives and the possibility of having null referential subjects. In taking these two parameters in all combinations, we get four distinct groups:

(i) [+null expletive], [+null referential]; e.g., Italian

(*Ciò) Piove. (Lui) Parla.

Indeed, changes in word order in Portuguese comes with a price. Costa writes that “word order variation is not pragmatically neutral; word-order alternations are related to different discursive functions” (2000a: 103).
(ii) [+null expletive], [-null referential]; e.g., Icelandic

\[ \textit{Rignir? *(Hann) talar.} \]

but \[ *(\textit{Thad}) \textit{Rignir.} \]

(iii) [-null expletive], [+null referential]; none

(iv) [-null expletive], [-null referential]; e.g., English

\[ *(\textit{It}) \textit{is raining.} *(\textit{He}) \textit{speaks.} \]

Although there are no languages identifiable by the third group above, Neapolitan may provide some insight into the success of this system of classification. While expletives in Neapolitan may have phonetic content (Ledgeway 2000: 77-78), they are not required to be overt. When they are non-null, pragmatics is involved; overt expletives normally have some causal reading associated with it (Adam Ledgeway, personal communication). So the sentence:

(5) \textit{Chillo chiove.}

'It is raining.'

is an appropriate response to a question such as “Why are you taking your umbrella?”. However, the omission of \textit{chillo} from (5) would be a general observation on the weather, with no discourse expectations attached.

The Neapolitan example above suggests that Rizzi’s framework might benefit from further refining. It may be helpful to include in this system, for instance, not merely whether languages allow a null expletive, as Icelandic and Neapolitan do, but also whether languages completely prohibit the overt realisation of expletives (as is the case in standard Italian).
An intended outcome of this study is to determine whether Rizzi's approach to the typology provides a sufficient account of the range of null subject usage we see across all languages. If we are going to talk about the licensing of null subjects as a (binary) parameter, which should not be automatically assumed (see Section 1.4.3), then what are we to make of (a) the situation with null expletives discussed above, and (b) languages like Icelandic and Neapolitan, both of which have mixed settings for different sorts of syntactic subjects?

Huang (1984) addresses the counterexamples used to argue against Taraldsen's proposal that null subjects are available only in languages with rich verbal morphology, by suggesting the following amendment, which is a mere stipulation: null subjects are permitted in languages that either have rich agreement or do not have any at all. Jaeggli and Safir pursue a similar path, stating that "null subjects are permitted in all and only languages with morphologically uniform inflectional paradigms", that is, paradigms that have "either only underived inflectional forms or only derived inflectional forms" (1989a: 30). This explanation would be compatible with what is seen in those Asian languages that freely allow omission of subjects but do not inflect their verbs. However, if we consider Afrikaans, which is not a null subject language, verbs do not inflect for person or number in this language, as is the case in Chinese. So both Huang's and Jaeggli and Safir's amendments to Taraldsen's proposal fall short of providing a complete and accurate account of the pro-drop phenomenon.

One of the goals of this research is to examine the links between inflectional paradigms, subject recoverability in the verbal morphology, and null subjects. It is hoped that in doing so, our discoveries will shed more light on why null subjects are permitted in certain languages and in certain contexts, while restricted in some, and still prohibited in others. While the primary focus will be on Romance varieties, data from non-Romance languages will be considered as well where appropriate. The aim is, then, to formulate a more accurate hypothesis, one that can be applied to all languages regardless of their family, of how pro, and perhaps null arguments in general, can be licensed.

---

5 See footnote 3.
1.2.2 Is it pro-drop?

It is also worth examining briefly what exactly does constitute an example of pro-drop, and what does not. The systematic omission of overt subject pronouns in certain environments is not sufficient evidence to assert that we are witnessing instances of pro-drop. Consider the English diary register, the type of written English found not only in personal diaries, but also in e-mails and sms text messages. We consistently see subject pronouns dropped at the beginning of sentences:

(6a) ( ) Went to the market today. ( ) Bought two apples.

On the surface, this appears to be identical to what is observed in the Romance languages. Compare the equivalent in Italian:

(6b) ( ) Sono andato al mercato oggi. ( ) Ho comprato due mele.

However, upon further inspection, that is as far as the similarities go. In other contexts where the subject pronoun in Italian could be left unexpressed, for example, in embedded clauses and with elements preposed to the subject pronoun (Haegeman 1990: 174), the equivalent construction in the English diary register is unattested and ungrammatical:

(7a) ( ) Don’t know if *(he) has seen me.
(7b) ( ) Non so se ( ) mi abbia visto.

(8a) When *(l) saw him, it was raining.
(8b) Quando ( ) l’ho visto, pioveva.
In fact, the differences noted above demonstrate that the omission of subject pronouns in the English diary register (and that of French as well) show similarities not to pro, but instead to wh-traces (Haegeman 1990: 174-175).

On a related note, the omission of subjects by young native speakers of English does not entail that the language starts out [+pro-drop] in the early stages of its acquisition and then transforms somehow to [-pro-drop] as the child’s exposure to the language increases. In fact, Rizzi (1994: 255) remarks that in English, this Early Null Subject seems to mirror the covert subject in the diary register with regard to distribution and syntactic restrictions. Based on her data and quantitative analysis, Valian (1994: 275) supports Rizzi’s claim by asserting that there is no evidence that American children ever have pro in their grammars.

It is possible that what is commonly considered an instance of something else may in fact be pro. In some varieties of Spanish spoken in the Caribbean, it is grammatical to place subjects between prepositions and infinitives (Suñer 1986: 194). Whereas the majority of Spanish speakers (including those whose first language is Castilian6) would find them completely unacceptable, sentences like (9) are perfectly fine to native speakers of this particular variety of Spanish:

(9) *Dije la verdad para tu no ser arrestado.*

‘I told the truth so that you would not get arrested.’

All Spanish speakers find the alternative without the subject pronoun grammatical, but note that (9) and (10) do not mean the same thing:

(10) *Dije la verdad para no ser arrestado.*

‘I told the truth in order (for me) not to get arrested.’

What is most striking about the construction in (9) is its similarity to the inflected infinitival construction in Portuguese. Compare the equivalents to the sentences above:

---

6 Throughout this entire work, we will use the term ‘Castilian’ to refer to standard European Peninsular Spanish.
The question that naturally arises, of course, is whether the subject pronoun placed directly before the infinitive assumes a position that was previously occupied by PRO or pro. If we go by traditional syntactic accounts, subjects of infinitives are PRO. However, Raposo (1987: 86) shows that the subject of the inflected infinitive in Portuguese is pro. This has broader ramifications on theories of binding and control, and also on the relationship between PRO and pro.

The value of these observations is that data telling us what is not pro, and even data that are ambiguous as to whether we are witnessing instances of pro, may be extremely beneficial to our study of the intricacies of the null subject.

1.2.3 How does a language “stop” being pro-drop?

Language is not a static entity; its constant use ensures that it is always changing and developing. It is natural to expect then that there will be instances in which certain properties of languages will evolve over time. Any work concentrating on null subjects and the pro-drop phenomenon would be wise to incorporate into its analysis, therefore, an investigation of the current situation in languages that straddle the line between parameter settings.

A quick reference to the current situation in some types of Latin American Spanish and Portuguese would be beneficial to the discussion here. A more detailed scrutiny of the status of subject pronouns in these varieties follows in subsequent chapters, but it is sufficient to mention here that, for example, an analysis of Miguel Falabella’s 1992 piece No Coração do Brasil shows that subjects occur overtly roughly three-quarters of the time (Duarte 1993: 111-112). If this is indeed indicative of the present state of the everyday spoken language, then can Brazilian Portuguese be unequivocally distinguished from its European counterpart, whose status as a null subject language is much less controversial? A similar situation holds in the Spanish
of the Dominican Republic (see Section 1.3.4); there is a notably higher instance of overt subject pronoun usage, to the extent that speakers of other varieties of Spanish would consider its ubiquity to be redundant, and perhaps even ungrammatical (Suñer 1986: 196). If speakers of incontestably [+pro-drop] varieties of Spanish find this variety of the language to be ungrammatical with respect to the use of subject pronouns, does it necessarily follow that the two varieties no longer share the same null subject parameter setting? What is clear from this data is that the pro-drop situation in these Latin American varieties is anything but obvious.

Furthermore, it is useful to make clear what is meant precisely when a language is described as being “in the process of becoming a non-null subject language”, as is asserted, to give an example, for Brazilian Portuguese by Kempchinsky (1984: 14). Does the switching of a parameter for a null subject language involve, for instance, restricting both the null expletive and the null referential subjects in the context of Rizzi’s (1982) framework, or will it suffice if only one of these, but not the other, is prohibited, for example, just null referential subjects? How can we incorporate the non-compulsory use of expletives occurring in some languages within the definition of a “null subject language”? These will be the sorts of questions that will have to be answered if we are to make substantial progress.

1.3 What the grammars say, and do not say

This section provides a brief survey of how grammars of various Romance languages treat this issue of the null subject. We start by examining the situation in Latin, and then in the languages for which there is no shortage of grammars, and then proceed finally to those languages for which relatively less has been published.

1.3.1 Latin

We start our investigation of how pro-drop is treated in the various Romance languages by examining how this issue was handled in these varieties’ mother language, Latin. As all six forms within any one of the classical Latin verbal
paradigms were distinct, there was no fear of ambiguity with respect to person and number, and so it is widely claimed that the use of subject pronouns occurred only to clarify or emphasise.

However, there is documentation that subject pronouns were used fairly often in the spoken language:

"The personal pronoun is not expressed in classical prose, unless it is emphatic, as for example, in contrasts. [...] The insertion of the pronoun without emphasis is very common in the comic poets, and seems to have been a colloquialism."

(Gildersleeve and Lodge 1895: 146).

Subject pronouns representing the persons directly involved in the discourse were employed often in Vulgar Latin:

"The personal pronouns came into more and more frequent use. Ego and tu are very common in Petronius."

(Grandgent 1908: 34)

So it has been recognised that there is a historical precedent for subject pronouns to be used in pro-drop languages that serve neither to emphasise nor to clarify person ambiguity on verbal desinences.

The Latin grammarians amongst themselves showed significant variation in how they treat what we now call pro-drop. On the one hand, Priscian, author of an enormous grammar, Institutiones grammaticae, written in 5th-6th century A.D. Constantinople, maintains that:

"si enim dicam 'scribo' vel 'scribis' in ipsa voce definivi etiam personam scribentis et ostendi; sin dicam 'scribit', incertum quis, donec addam vel nomen vel pronomen."

(Keil 1870: 578)

7 "Indeed if I should say 'scribo' ['I write'] or 'scribis' ['you write'], in that word I will have marked out the person; whereas if I were to write 'scribit' ['? writes'], it would be uncertain who [the subject is], unless I were to add a noun or a pronoun."
This suggests a recognition of differences in subject pronoun usage based on, notably, grammatical person; there is a clear distinction made between verbs whose subjects are discourse participants (and hence, more salient and immediately identified), and verbs with third person subjects. On the other hand, Donatus, a famous Latin grammar from the 4th century A.D., does not discuss subject pronoun usage in either the pronouns or verbs sections of his *Ars Minor* (Keil 1870: 357-362; 379-385). This is particularly noticeable in light of his thorough treatment of all the other aspects of Latin pronouns and verbs, such as inflections, irregularities, and exceptions. In the absence of any other comment on the matter, our observation that subject pronouns are never included next to any of the verbs in that section of Donatus' grammar may facilitate the interpretation that Latin subject pronouns were indeed considered truly optional, or perhaps even ungrammatical unless used in specifically dictated circumstances. In any case, there was no wide agreement on the employment of subject pronouns in Latin, and this ambiguity continues into its modern daughter languages.

### 1.3.2 Italian

Arguably the most often cited example of a non pro-drop situation in a pro-drop language comes from Italian and the requirement of the personal pronoun *tu* in subjunctive clauses with a second person singular subject. In the present subjunctive, all three singular persons share the same verbal form; indeed, the presence or absence of a subject pronoun has identifiable ramifications for the interpretation of the clause:

> "Se il verbo non viene preceduto dal pronomne 'tu', esso viene interpretato, a seconda del contesto, come una prima o come una terza persona."

(Cordin and Calabrese 1988: 540)

(13a) and (13b) below illustrate this:

---

8 "If the verb is not preceded by the pronoun *tu*, it is interpreted, according to the context, as a first or third person."
Cordin and Calabrese go on to demonstrate the case with the imperfect subjunctive, where only the first and second person singular share the same grammatical form. The compulsory use of *tu* for a second person subject is also in effect here:

(14) *Credevano che *(tu) andassi con loro.*

‘They believed that you were going with them.’

(adopted from Cordin and Calabrese (1988: 541))

What is not certain, though, is whether this restriction still holds if the discourse or co-reference makes clear that *tu* is the intended subject of the embedded clause. Consider the verb *pentirsi* (to repent); it is a particularly good test case, as it is a verb that does not exist non-pronominally and therefore, there is no danger of interpreting this verb as being transitive. In other words, *mi/ti/si* must be interpreted as an inherent reflexive clitic that matches its associated subject. Lepschy and Lepschy state that:

“Even with pronominal verbs, where the form without subject pronoun would not be ambiguous, *vogliono che tu ti penta* is used, rather than *vogliono che ti penta* ‘they want you to repent’ [.....] [but] this does not mean that examples without the pronoun are not to be found.”

(Lepschy and Lepschy 1988: 143, 160)
So it is acknowledged that an otherwise absolute restriction does show some weakening when the subject is recoverable through a combination of the syntactic clitic and the semantics of the pronominal verb.

1.3.3 Portuguese

Most grammars of the Portuguese language handle the pro-drop phenomenon lightly by appealing to the rich verbal inflection of its paradigms. Mateus, Brito, Duarte, and Faria mention as one of the typical properties of the subject in Portuguese:

"Dadas as características da flexão verbal em português, quando o [sujeito] é um pronome não enfático tem, geralmente, uma realização nula."

(Mateus et al. 1989: 162)

What is particularly striking is the example they use right after the description above:

"Soube que passaste no exame. Parabens!" é mais natural do que 'Eu soube que tu passaste no exame. Parabens!"  

(Mateus et al. 1989: 162)

As soube could be either a first person or a third person singular, it should be possible for the subject of the sentence to be ele. Of course, the use of the interjection Parabens! following the sentence highly encourages the interpretation of the subject of soube to be a first person rather than a third person. However, this is an example of discourse analysis and pragmatics at work. From a syntactic point of view, there should be nothing to prevent the listener from interpreting the example above as, "He

9 "Given the characteristics of verbal inflection in Portuguese, when the subject is a non-emphatic pronoun, it has, in general, a null realisation."
10 "(I) found out that (you) passed the exam. Congratulations!" is more natural than 'I found out that you passed the exam. Congratulations!'"
found out that you passed the exam. Congratulations!”, unless there is a formal prohibition at work here that has not been explicitly detailed (e.g., an automatic default interpretation of a first person subject in these situations as a result of salience of the speaker or pragmatic and discourse considerations).

Cunha and Cintra’s description of the Portuguese language brings to the forefront two additional observations. First, there is the common treatment of null subjects that is ubiquitous in Romance grammars:

“Os pronomes sujeitos eu, tu, ele (ela), nós, vós, eles (elas) são normalmente omitidos em português, porque as desinências verbais bastam, de regra, para indicar a pessoa a que se refere o predicado, bem como o número gramatical (singular ou plural) dessa pessoa. [...] Emprega-se o pronome sujeito: (a) quando se deseja, enfaticamente, chamar a atenção para a pessoa do sujeito; [...] (b) para opor duas pessoas diferentes; [...] (c) quando a forma verbal é comum à primeira e à terceira pessoa do singular e, por isso, se torna necessário evitar o equívoco.”

(Cunha and Cintra 2000: 284-285)

However, under a later section entitled “Verbal Agreement”, we see the following statement and example:

“A concordância evita a repetição do sujeito, que pode ser indicado pela flexão verbal a ele ajustada: “Eu acabei por adormecer no regaço de minha tia. Quando acordei, já era tarde, não vi meu pai.”

(Cunha and Cintra 2000: 494)

While the example is used to show why it is not necessary to repeat eu before acordei and vi, the grammar does not explain why it is even necessary in the first place to use

11 “The subject pronouns eu, tu, ele (ela), nós, vós, eles (elas) are normally omitted in Portuguese because, as a rule, the verbal desinences suffice to indicate the person to which the predicate refers, as well as the grammatical number (singular or plural) of that person. [...] The subject pronoun is used (a) when it is desired to call emphatic attention to the person of the subject; [...] (b) to oppose two different persons; [...] (c) when the verbal form is common to the first person and third person singular, and as a result, it becomes necessary to avoid ambiguity.”

12 “Agreement allows avoidance of the repetition of the subject, which can be indicated by the verbal inflection linked to the subject: I ended up falling asleep in my aunt’s lap. When (I) woke up, it was already late, I didn’t see my father.”
an overt subject pronoun for *acabei*, as the verbal form is unambiguously a first
person singular. The question is, therefore, whether the presence of *eu* renders the
sentence grammatical when it otherwise would not be, or whether its overt use has to
do instead with focus, topic, or discourse considerations.

The second observation to note from this grammar is the increasing use of an
overt expletive in colloquial Portuguese:

“Na linguagem popular ou popularizante de Portugal aparece por vezes um pronomene
ele expletivo, que funciona como sujeito grammatical de um verbo impessoal, à
semelhança do francês il (il y a): Ele haveria no mundo nada mais acertado.”

(Cunha and Cintra 2000: 284)

What is evident in this example is Portuguese’s similiarity to Neapolitan
(discussed in Section 1.2.1 above), and the problems that this optional overt expletive
brings to Rizzi’s framework on language typology and null subjects. This is further
evidence, then, of the need to re-evaluate how the *pro*-drop phenomenon is
characterised.

1.3.4 Spanish

As there are 350 million speakers of Spanish, most of whom doing so natively
(Mar-Molinero 1997: 4), it will come as no surprise that there is tremendous variation
in the usage of subject pronouns in the Hispanophone world. In Castilian, where
overt subject pronouns are relatively rare compared to the same in Latin American
varieties, we find that there is no clear correlation between verbal ambiguity and
subject pronoun usage:

“[E]l morfema de persona incluido en el verbo distinguee ya cuál de las tres funciona
como sujeto grammatical, y así no resulta muy necesaria la presencia de un
sustantivo personal para señalar un sujeto explicito: en canto, cantas, canta, están ya

---

13 In the popular, or increasing popular, language of Portugal, there appears at times an expletive
pronoun *ele*, which functions as a grammatical subject of an impersonal verb, similarly to the French *il*
(*il y a*): There was nothing in the world more striking.”
expresas como sujeto las personas primera, segunda, y tercera, respectivamente. No obstante, es frecuente la aparición de un personal en esa función de sujeto explícito, y no solo en los casos de coincidencia fonética de las formas verbales (como cantaba, cantaria, cante, en que no se distingüe la primera de la tercera persona), ni en el caso de la tercera persona (donde la distinción de géneros del personal puede aportar mayor precisión acerca de la referencia concreta al sujeto). También pueden aparecer yo y tú, aunque su referencia personal es evidente e inequívoca en cada acto de habla.\textsuperscript{14}

(Alarcos Llorach 1994: 73)

So personal pronouns may be overt both in those circumstances where the verbal morphology is ambiguous with regard to person, as well as in those situations (i.e., the first and second persons) where there is no difficulty in interpreting the intended subject from the verbal ending.

There is also recognition that the decision to use or omit overt subject pronouns is not one dictated completely by chance and whim:

"[E]l pronombre sujeto en español no está siempre expresado explícitamente en el contexto. Esto no significa en absoluto, sin embargo, que se trate de un sistema caprichoso en el que los pronombres pueden estar o no estar por pura casualidad. Si bien es cierto que el verbo en sí mismo, diferentemente de lo que encontramos en otros idiomas, ya contiene las marcas personales (incluso en la lengua hablada), hay casos en los que la presencia del pronombre personal sujeto se hace imprescindible para una correcta comprensión."\textsuperscript{15}

(Matte Bon 1992: 246)

\textsuperscript{14} "The person morpheme included on the verb already distinguishes which of the three persons functions as the grammatical subject, and therefore the presence of a subject personal pronoun is not very necessary to signal the explicit subject: in canto, cantas, canta, the first, second, and third persons, respectively, are already expressed as the subject. However, the presence of a personal pronoun is frequent in the role of explicit subject, and not only in those cases of phonetic coincidence of verbal forms (as in cantaba, cantaria, cante, in which the first and third person cannot be distinguished), nor in the case of the third person (where the distinction of gender of the personal pronoun can be better specified through concrete reference to the subject). In addition, yo and tú can appear, even though their personal reference is evident and unmistakable in each speech act."

\textsuperscript{15} "[T]he subject pronoun in Spanish is not always expressed explicitly in the context. This absolutely does not mean, however, that the system is unpredictable in that pronouns can appear or not appear randomly. Although it is certain that the verb itself, differently from that found in other languages, already contains markers for person (including in the spoken language), there are cases in which the presence of a personal subject pronoun is indispensable for a correct interpretation."
So the implication that there is indeed some sort of formal system of subject pronoun usage, and that the presence or absence of overt subject pronouns is not "optional", motivates the current research to identify such a system in great detail.

It should be pointed out as well that the use of subject pronouns in several varieties of Latin American Spanish is frequent, and this usage is often described as being "redundant" (cf. Lipski 1994: 241). Furthermore, there is debate as to the connection between the loss of final consonants on verbal endings, the ensuing loss of distinction in the verbal paradigm, and the higher incidence of subject pronoun use, as in the case of deletion of word-final /s/ in Puerto Rican Spanish (Lipski 1994: 334-335). One of our goals is to determine whether these phenomena are truly linked, and if so, whether one triggers the other.

1.3.5 French

There is certainly no lack of sources describing French as a non-null subject language and insisting upon the compulsory use of subject pronouns. More interesting is the documentation of some current changes occurring in the modern-day language that may cause some reconsiderations of how French is viewed. Bonnard mentions that:

"tu est éliéd en franç;ais familier: T'as raison...ii et ils sont prononcés couramment [i] devant consonne."\(^{16}\)

(Bonnard 1990: 184)

The phonetic weakening of these three subject pronouns is one indication that they may not act entirely like the tonic subject pronouns in other, non-null subject languages.

Another pattern worth investigating is the juxtaposition of disjunctive pronouns and their unstressed counterparts:

\(^{16}\) "Tu is elided in familiar French: You're right...il and ils are currently pronounced [i] in front of a consonant."
"Nowadays, moi, toi, and stressed nous, vous cannot function as the grammatical subject of a verb but must be accompanied by the corresponding unstressed pronoun, je, etc: moi je crois que... vous, vous êtes français mais nous, nous sommes anglais."

(Price 1998: 145)

This may suggest that what we normally consider to be the French subject pronouns should be better characterised as part of the verbal construct (i.e., that perhaps moi may in fact be a subject pronoun in a French pro-drop framework, and that the sequence je crois constitutes the first person singular verbal form of croire\textsuperscript{17}).

Finally, there have been arguments put forward in favour of treating French as a null subject language (cf. Roberge 1986). Therefore, there is reason not to assume automatically that French is unequivocally a non-null subject language without a thorough inspection of phenomena such as those discussed above.

1.3.6 Catalan

The pro-drop situation is not entirely clear in Catalan. The three grammars discussed here make three distinct comments on this topic. Hualde mentions that:

“subject pronouns may be left unexpressed. In fact, they are usually left unexpressed, except for emphasis or contrast. This is regardless of whether person and number information is transparently encoded in the verb. Null subject pronouns are also possible if the verb is in one of those tenses, such as the imperfect and the conditional, where the endings of the first and third person singular are identical:

\textit{En Joan va dir que sabia francès.}  
John, said that I/he\textsubscript{u} knew French.”

(Hualde 1992: 167)

\textsuperscript{17} See Harris (1982; 1988) for more on this possibility.
There is no explicit mention of which methods one would use in the example above in order to ensure the exclusive interpretation of the subject of *sabia* as either *Joan*, another third person, or the speaker.

Wheeler et al. address this issue from a different perspective:

"There are some cases, however, in which the verbal forms for first- and third-person singular coincide, namely in the imperfect and conditional tenses (jo cantava – ell/ella/vostè cantava; jo dormia – ell/ella/vostè dormia; jo dormiria – ell/ella/vostè dormiria) and the context may require [bold PSM] the pronouns to be used to avoid confusion".

(Wheeler et al. 1999: 164)

So one grammar tells us that, with verbal forms ambiguous with respect to person, null subject pronouns are possible, while another one tells us that in certain situations, the use of pronouns may be obligatory.

Finally, Jané takes a much simpler, and arguably more nonchalant, approach to this question:

"*Els pronoms personals forts* (jo, tu, ell, ella, nosaltres, vosaltres, ells, elles, nós, vós, i voste) exerceixen la funció de subjecte del verb, és a dir, d’aquell qui la realitza, i el seu ús no ofereix particularitats dignes d’esment."\(^{18}\)

(Jané 1979: 127)

That there is no consensus among these three selected references is strong testimony to both the justification of pursuing more in-depth studies on *pro*-drop in Catalan, and also to the necessity of a clearer understanding of the null subject.

---

\(^{18}\) "The strong personal pronouns (jo, tu, ell, ella, nosaltres, vosaltres, ells, elles, nós, vós, i voste) serve as subject of the verb, that is to say, of that which realises it, and their use does not offer particularities worthy of attention."
1.3.7 Galician

This survey of grammars of the Galician language will bring critical attention to two main points. We start first by excerpting one grammar's treatment of the null subject:

“Ainda que na maior parte dos casos a desinencia verbal abonda para indica-la persoa e o número a que se refire o predicado, a forma do pronome suxeito vén con frecuencia expresa, se ben a súa presencia nunca é obrigada.”19

(Álvarez, Monteagudo, and Regueira 1986: 166)

The use of nunca causes us to consider whether the presence of subject pronouns is indeed never compulsory. Such a statement implies that overt subjects are truly optional, and have no effect whatsoever on the meaning of the utterance, although we strongly doubt this.

That being said, consider the following statement, from the introduction of the same grammar:

“Poderanse atopar, con toda seguridade, lagoas ou omissoes, atribuíbles ó descoñecemento por falta de suficientes estudios previos ou por deficiencias da nosa formación ou información. [...] O feito de que non se mencionen non quere dicir que non existan e que non poidan ser correctos.”20

(Álvarez et al. 1986: 8)

The authors admit here that not every aspect of the Galician language has been studied or documented, and as such grammars will not be complete or exhaustive. That certain facts and phenomena are not discussed in these grammars does not imply

---

19 “While in the majority of cases the verbal desinence suffices to indicate the person and number to which the predicate refers, the subject pronoun is frequently expressed, even though its presence is never obligatory.”

20 “It is very sure that holes and omissions will be found; this is attributable to a lack of knowledge, either because of a lack of sufficient studies in the past or because of deficiencies in our education or information. The fact that they are not mentioned does not mean that they do not exist and that they cannot be amended.”
that they do not exist or that they cannot be questioned. This is further motivation for a rigorous study of pro-drop, particularly in those Romance varieties for which relatively little has been examined, and for which there is a great need to delve further into this topic in order to attain a better understanding of it.

The second main point can be demonstrated by considering the discussion in another grammar, this one written in Spanish and written for an audience of non-native speakers of Galician. Carré Alvarellos gives several examples of usage of Galician personal pronouns, with their Spanish equivalents. Two of these, occurring very closely to each other, are particularly notable:

(15a) “O non sei.” (Galician)
(15b) “No lo sé.” (Spanish)

‘I don’t know.’

(16a) “Eu sei os que foron.” (Galician)
(16b) “Yo sé los que fueron”. (Spanish) (Carré Alvarellos 1967: 65)

‘I know those who went.’

Not only is it not made clear why it is permissible (or perhaps obligatory) to use the personal pronoun eu in (16a), or why (15a) does not require eu, but also there is no mention of how to use personal pronouns at all in this section entitled “Pronombres Personales”. This leads one to question whether there is an inherent assumption by the author that, since speakers of Spanish are familiar with their own usage of pro-drop, the system in Spanish is automatically applicable in every way to the Galician language. In other words, this presumption that rules on subject pronoun usage are equally valid to any Romance language assumes that there is a unitary pro-drop phenomenon in this family of languages. This bias must be scrutinised for its accuracy.
1.3.8 Occitan

A treatment of personal pronouns in a grammar of Occitan written in French brings up questions in both of those languages:

"Le verbe provençal se conjugue, nous l'avons vu, sans pronom personnel sujet. Mais il admet l'emploi d'un pronom (ieu, tu, éu, elo, nautre, vautre, éil) dit pronom d'insistance qui sert à renforcer le sujet contenu dans le verbe:

<table>
<thead>
<tr>
<th>Occitan</th>
<th>French</th>
</tr>
</thead>
<tbody>
<tr>
<td>ieu parle</td>
<td>moi, je parle</td>
</tr>
<tr>
<td>tu escoutes</td>
<td>toi, tu écoutes</td>
</tr>
<tr>
<td>éu legis</td>
<td>lui, il lit</td>
</tr>
</tbody>
</table>

(Bayle 1982: 171)

The first issue is whether the only function of the overt subject pronoun is really to reinforce the subject expressed by the verbal ending. There is brief treatment of this issue elsewhere in the grammar, only a simple claim that verbs in Occitan are conjugated without personal subject pronouns (cf. Bayle 1982: 75, 101).

The second topic to be queried lies in the French translations of the Occitan examples cited above. If emphatic ieu parle is to be translated as an emphatic moi, je parle in French, then does this imply that the French disjunctive pronoun moi is the same as Occitan ieu? Would escoutes in Occitan, then, be formally equivalent to the French sequence tu écoutes as solely verbal forms, with Occitan tu and French toi acting as subject pronouns? Examples such as this fuel the need for a full discussion of the situation in French as presented in Section 1.3.5 above.

1.3.9 Sardinian

Sardinian, traditionally divided between its southern (Campidanese) and northern (Logudorese-Nuorese) dialects (Jones 1997: 376), differs from most of the

---

21 "The Provençal [Occitan] verb is conjugated, as we have seen, without a personal subject pronoun. However, it does allow the use of a pronoun (ieu, tu, éu, elo, nautre, vautre, éil), called a pronoun of insistence, which serves to reinforce the subject contained in the verb: I speak, you listen, he reads, etc."
other languages we have investigated in this section in that it suffers from a lack of standardisation. Jones writes that:

“there is no single dialect which is recognised as a standard form of the language and there is no standard orthography. Sardinian has not enjoyed any official status since the Middle Ages. [...] There are few written texts in contemporary Sardinian (apart from dialect poetry, whose language tends to be rather artificial), the language being used mainly for informal oral communication.”

(Jones 1988: 314)

It should not come as a surprise, then, that Sardinian grammars show considerable variation in their treatment of pro-drop. A rather thorough discussion of the null subject phenomenon is presented by Jones, who suggests among other things that in Sardinian, all verbal paradigms have six distinct endings for all six persons, thereby allowing for the possibility of treating the verbal ending as an equivalent to the preposed subject pronoun seen in non pro-drop languages (1993: 15). Compare this to a grammar written in Italian by Blasco Ferrer, who does not directly mention the usage of subject pronouns with verbs, in either of his sections on personal pronouns or verbs (1994: 131-155). The closest he comes to treating subject pronouns in the context of the Sardinian verbal paradigm is his chart implying direct links between the subject pronouns and verbal desinences, although it is intriguing that the subject pronouns that are listed are those of standard Italian and not Sardinian:

<table>
<thead>
<tr>
<th>1ª persona singolare</th>
<th>2ª</th>
<th>3ª</th>
<th>Logudorese</th>
<th>Campidanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>'io'</td>
<td>'tu'</td>
<td>'lui'</td>
<td>-o, a, e</td>
<td>-u, a, i</td>
</tr>
<tr>
<td></td>
<td>'tu'</td>
<td>'lui'</td>
<td>-s</td>
<td>-s, -st</td>
</tr>
<tr>
<td>1ª persona plurale</td>
<td>'noi'</td>
<td>'voi'</td>
<td>-t</td>
<td>-t</td>
</tr>
<tr>
<td>2ª</td>
<td>'noi'</td>
<td>'voi'</td>
<td>-mus</td>
<td>-us</td>
</tr>
<tr>
<td>3ª</td>
<td>'noi'</td>
<td>'voi'</td>
<td>-des</td>
<td>-is</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-n</td>
<td>-nt</td>
</tr>
</tbody>
</table>

(Blasco Ferrer 1994: 153)

The chart above visually echoes Jones’ proposition of treating verbal desinences as the equivalent of preverbal overt subject pronouns. There are a number of pertinent questions that arise from this review of Sardinian. How accurate is it to
view verbal desinences as alternate versions of lexical subject pronouns? Is this something that can only be done in verbal paradigms lacking any syncretism (as is the case with Sardinian here, as opposed to the other Romance varieties we have considered), or can it be applied as well in situations where there is verbal ambiguity? Is this supposed immediate link between subject pronouns and verbal morphology valid cross-linguistically, or even in Sardinian, for that matter? It is hoped that we will be able to touch upon some of these issues in the present research.

1.4 Goals of this work

One of the aims of doing research on this topic is that a greater and more detailed knowledge of how pro-drop operates will help shed light on questions both central and peripheral to the null subject phenomenon. Not only does this work attempt to achieve a more accurate description of how subject pronouns are used and dropped, but also it seeks to address how this new information is relevant to other topics in linguistics as well. This section provides a brief synopsis of how a more thorough understanding of pro-drop would be a beneficial contribution to this field of study.

1.4.1 Are subject pronouns always optional?

We have already devoted much attention to the widespread assumption that subject pronouns in the Romance languages are never obligatory, citing statements present in a wide range of sources, in several languages, to reflect the omnipresence of this deep-rooted belief. We do not hide our scepticism of this claim. Even a casual consideration of the examples presented here should be convincing enough evidence of the imperfections of this assertion. Our task, then, is to identify the precise linguistic conditions that affect the licensing of the null subject, paying particular attention to the effects of various morphosyntactic environments and the accessibility of antecedents used for co-reference.
1.4.2 Is pro-drop uniform within Romance?

Can we accurately propose one model that is able to account for all the pro-drop phenomena attested in the Romance languages? Is there really just one single set of rules that can accurately dictate where subject pronouns must be used in this language family? We have already seen a number of linguists assuming this, but what is needed is a thorough and systematic investigation, such as the one that we hope to present in the current work. It is worth keeping in mind to what extent our findings are homogeneous in the different Romance varieties we will be considering.

It may well be difficult to argue that pro-drop in Romance is a unitary phenomenon if we acknowledge that subjects are made overt as a result of any combination of syntactic, phonological, or pragmatic reasons. In addition, if we accept arguments in favour of treating French as a null subject language (as proposed by, among others, Roberge (1986: 72) and Pierce (1994: 319)), then this situation does not get any less tricky. The benefit of addressing this question is that it may prevent in the future general statements such as (3) and (4) that are under specific in explaining null subject usage.

1.4.3 A null subject “parameter”?

It is normally taken for granted that the presence of null subjects is best explained as a parameter. Is this concept of a binary switch an accurate assessment of the situation? Perhaps after further review, it would be more appropriate to describe this parameter instead as a balancing scale, an extension of Valian’s metaphor on how children develop the part of their grammar that interprets null subjects (1984: 282). To mention another figurative assessment of the null subject “parameter”, it could be that Ross’ (1982) description of “hot” and “cold” languages (Huang 1984: 531) is more fitting, allowing for the possibility of “warm” languages that are pro-drop only in most but not all environments, and “cool” languages that infrequently do so. It may be that a cline is the optimal approach in characterising to what extent a language is pro-drop; this could avoid troublesome either-or classifications of, say, German as a
non-null subject language (although some expletives may be covert) and Italian as a null subject language (despite the requirement of tu in some syntactic environments). Furthermore, the idea of a cline, which would necessarily preclude the formulation of pro-drop as a parameter, would be compatible with a mixed-system like Hebrew, where null subjects are permitted in past and future tenses but not in the present (Borer 1989).

The speculation above does not assume that there has already been a decision made to abandon the parameter model. However, the discussion in this chapter has made it clear that there is much room for refinement in the treatment of the pro-drop phenomenon as a single parameter. It is quite possible that there are several individual parameters, each responsible for one certain aspect of null subject usage in a specific linguistic environment, that can accurately account for the wide range of pro-drop activity we observe in all languages when these numerous parameters are set in various combinations. It is equally possible that parameters are not so useful, or at least, much less informative than other methods in representing the null subject phenomenon. We now direct our attention to answering this and the other questions posed in this introduction.
Chapter 2: Identifying what does not affect the licensing of pro

2.1 Investigating what is not relevant to this study

In Section 1.2.2, we looked at instances that superficially appeared to be cases of pro-drop, but turned out actually to be a result of other linguistic phenomena. We also examined a grammatical construction in Caribbean Spanish that draws into question whether the entity under consideration is PRO or pro.

These investigations proved to be quite useful because they helped us to identify the domains in which pro exists. By having a greater understanding of both the areas in which we do not encounter pro, and also those areas where we are not completely sure whether it is pro that we are dealing with, we can devote more attention to the investigation of the behaviour of pro itself. This will allow us to focus most of our energy on examining specifically those environments where pro does operate.

Using a similar line of reasoning, we can then accept that there is great benefit in clarifying those issues that do not directly affect the licensing of pro. Once we are able to establish which factors are not important in our attempts at understanding how and where the null subject functions, we can devote our primary efforts to identifying and then deconstructing those factors that are influential in licensing pro.

Therefore, it seems prudent to engage first in some discussion on matters that are not crucial in determining the availability of the null subject, before scrutinising the issues pertinent to the licensing of pro in the chapters that follow. This chapter is concerned with examining those factors that are not directly relevant to the licensing of pro. We begin by inspecting some linguistic attributes, grammatical relations, thematic roles, and animacy, that we may expect a priori to be significant in determining the distribution of pro, but that turn out in fact not to be so critical in our investigation of the null subject. This is then followed by a brief review of the historical development of null subjects in French, and how its lessons may be an applicable precedent to our understanding of subject pronoun usage in other present-day Romance varieties.
2.2 Grammatical relations and thematic roles

2.2.1 Canonical relationships

We begin this section by examining grammatical relations, thematic roles, and the connection between the two. “Grammatical relations” is a broad term covering almost any type of association between “a main verb or predicate and its dependent arguments” (Croft 1990: 101). It is sometimes used synonymously with the expression “syntactic relations”, since the focus is on how a noun phrase and its predicate are linked together (Comrie 1989: 65). As identified by, among others, Palmer (1994: 241-242) and Blake (1990: 1), the grammatical relations normally taken under consideration are subject, direct object, indirect object, and oblique.

Thematic roles express how an entity participates in the event expressed in the sentence. There is no widespread agreement in the literature as to how many theta roles there are, or as to the exact boundaries between distinct but similar theta roles, but for our purposes, we can name the five commonly recognised theta roles filled by animate noun phrases, and the sort of involvement each role expresses in the sentence. This list, of course, is not meant to be exhaustive:

(i) AGENT – the doer or actor of an event
(ii) THEME – the entity which is changed or undergoes an action

1 Traditional grammars and dictionaries tend to define “oblique” as any noun case other than the nominative and the vocative. By this definition, the “direct object” and the “indirect object” would be considered as an instance of “oblique”. More contemporary references tend to represent “oblique” as any entity other than the subject, direct object, and indirect object, a sort of default group for anything that does not fit into one of the three commonly acknowledged grammatical relations. It should be noted that many typologies make separate references to direct/indirect objects and to obliques (Comrie 1989; Croft 1990; Song 2001, *inter alia*), thereby implying that the conservative definition of “oblique” is not intended. We follow their lead here and consider “oblique” to be anything other than a subject, direct object, or indirect object.

2 There is abundant literature covering the debate on the number and types of thematic roles in existence. Saeed (2003: 141) is a good place to start for those interested in the finer points of the discussion.

3 Foley and Van Valin (1984: 27ff.) use the term ACTOR for this thematic role, while some scholars use the terms ACTOR and AGENT indistinguishably, such as Haegeman (1994: 49) and Radford (1988: 373). Furthermore, Saeed (2003: 84) makes the distinction that AGENT, but not ACTOR, voluntarily initiates the performance of an action. We choose to use the more common term AGENT only throughout this work for this thematic role.
(iii) EXPERIENCER – the entity which undergoes an emotional or psychological experience

(iv) BENEFACTIVE/MALEFACTIVE – the person on whose behalf/against whom something is done

(v) RECIPIENT – the receiver of an action or item

Systematic correspondences between certain grammatical relations and particular thematic roles are often drawn. This has been recognised as early as the fifth century B.C., when the Sanskrit grammarian Panini classified subjects as doers of an action and objects as those people or things that undergo the action (Kearns 2000: 188), thereby making an explicit connection between an NP’s grammatical function and its expected semantic function. An initial consideration of this tendency might lead us to believe that it is a reasonable assumption to link a specific grammatical relation to some natural counterpart among the thematic roles available. There does seem to be an intuitive mapping between a grammatical subject, that which tends to be the initiator of some action expressed by the sentence, and the thematic role AGENT. The same can likewise be said about a grammatical direct object, that which is conventionally the entity undergoing the action described in the sentence, and the thematic role THEME. This can be extended to other grammatical relations and thematic roles as well: for example, one would expect a natural correlation between a grammatical indirect object and the thematic role RECIPIENT, as both are commonly associated with being only peripherally involved in the principal action holding between the subject/AGENT and the direct object/THEME.

However, a more discerning investigation of this situation will point out the weaknesses of this idea of an inherent correspondence between grammatical relations and thematic roles. One strong counterargument is the existence of the passive construction. When the passive voice is used, the THEME, and in some languages, other thematic relations as well, such as RECIPIENT in English (e.g., ‘He was given a present’), is promoted to grammatical subject, either for the purposes of emphasis, discourse salience, or defocusing of the agent, and AGENT is demoted to an oblique adjunct. So passive sentences bring about a mapping of AGENT and THEME with oblique and subject, respectively. This is an apparent mismatching of the supposed natural ties between grammatical relations and thematic roles. What is even more
convincing proof against the argument for expected links is that passive constructions are not at all uncommon; they are widely attested cross-linguistically, and so there can be no asserting that this method of expression is a relatively rare anomaly, albeit marked.

Another persuasive argument against the claim of inherent connections between grammatical relations and thematic roles is the great variation in expressing grammatical functions in the world’s languages. In English, for instance, subjects of transitive and intransitive verbs are similar in that they both occupy the same position in the sentence and trigger agreement with the verb with regard to person and number. However, they may differ with respect to thematic role; the subject of a transitive verb is likely to be AGENT, whereas the subject for an intransitive verb may be, for instance, EXPERIENCER (e.g., ‘I thrived on the adrenaline rush’) or THEME (e.g., ‘The ship sank’). But the subject of an intransitive (i.e., unergative) can also be AGENT (e.g., ‘to smoke’), in the same way that the subject of a transitive might also be EXPERIENCER (e.g., ‘John feels the pain’).

Languages with ergative systems, on the other hand, do not operate in the same manner. For example, in Dyirbal, a native Australian language, the subject of an intransitive verb is grouped with a transitive direct object, as both carry an absolutive case marking. Subjects of transitive verbs, on the other hand, take a different case marker, the ergative (Dixon 1972: 59). So transitive direct objects, which are likely to have the thematic role THEME, are treated syntactically like subjects of intransitive verbs (be they unergatives or unaccusatives), which tend not to be THEME.

These explanations above should be sufficient justification for considering grammatical relations and thematic roles to be two independent factors. We have seen evidence, both within the system of one language, and also by comparison between different languages, of the separate qualities of grammatical relations and thematic roles. In Section 2.2.4 below, we investigate whether either has any direct effect on the licensing of pro.
2.2.2 Switch reference

In this section, we look at switch reference, how it is treated in the languages we are studying, and to what extent it is relevant to the present research. While switch reference encompasses a wide range of circumstances, the meaning of switch reference that we are particularly interested in is that used to describe a situation where the subject of a sentence is not the same as the subject of the previous sentence. A simple instance of switch reference is illustrated below in (1), where the subject of the second sentence is not co-referential with that of the first. (2) is an example of same reference, the subjects of the two sentences being identical to each other:

(1) John went with Jane to the hospital yesterday. She wasn’t feeling well, so he drove her there in his car.

(2) John went with Jane to the hospital yesterday. He wasn’t feeling well, so she offered to drive him there in her car.

Switch referencing is not treated identically in the world’s languages. We see above that in English, the use of an overt subject pronoun to refer to the intended antecedent of the subject of the second sentence is sufficient means to license a new subject. In Japanese, however, this is done with the use of postpositional markers *wa* and *ga* (Takeuchi 1999: 139). Still other languages resort to alternate means, such as inflexional affixes on the verb. This is the case with Dyirbal, where the verbal suffix *gay* is used to signal a switch, whereby a previously mentioned THEME is the current AGENT (Silverstein 1976: 154-5). Finally, more rarely do languages indicate switch reference via independent morphemes not bound on the verb.

Those Romance languages that permit the null subject in certain situations treat these phenomena of same and switch references in a different manner. Normal practice for indicating same reference requires a null subject at the beginning of the second sentence co-referential with the subject of the previous sentence, as the European Portuguese example below demonstrates:
(3a)  O João, foi com a Maria para o hospital hoje. (pro), Sentia-se doente.
   ‘John, went with Mary to the hospital today. (pro = John), was feeling ill.’

In some Romance varieties, it is considered awkward and redundant to use an overt subject pronoun co-referential with the subject of the previous sentence; certain speakers of these varieties may even find (3b) ungrammatical:

(3b)  ?? O João, foi com a Maria para o hospital hoje. Ele, se sentia doente.

Indeed, one of the goals of the present research is to identify which Romance varieties resort to pro to indicate same reference only, and which are more liberal by allowing it to be co-referential as well with an entity other than the previous subject.

As for switch reference, the situation is not as straightforward. All Romance varieties permit the use of an overt subject pronoun to indicate switch reference, paralleling what we saw in English in sentences (1) and (2). The Spanish sentence in (4) shows this property, and it is representative of how switch reference is indicated in Romance:

(4)  Juan fue con Maria al hospital esta mañana. Ella no estaba bien.
   ‘John went with Mary to the hospital this morning. She was not feeling well.’

However, while certain Romance varieties require the use of an overt pronoun as the subject of the second sentence if it is not co-referential with that of the first sentence, others will tolerate its null counterpart, subject to particular conditions being met. This may include morphological agreement on the verb of the second sentence that would preclude co-referencing with the subject of the first sentence (e.g., the verb of the second sentence is marked as plural, but the subject of the first sentence is singular), or some indication of gender in the predicate of the second sentence, be it through semantic or syntactic means, that would make it logically impossible for its subject to be interpreted as co-referential with that of the first sentence (e.g., the subject of the first sentence is masculine, the predicate of the second sentence indicates that its subject is feminine). To illustrate this more clearly, Castilian Spanish
will allow the absence of an overt subject in cases of switch reference if the verbal morphology or semantics can identify the antecedent in the previous sentence:

(5)  \[ \text{Juan va a quedar con los vecinos, hoy. (pro), Lo odian.} \]

'John is meeting the neighbours today. (pro = The neighbours) hate him.'

(6)  \[ \text{Pablo acompaña a Laura, a ver los resultados del examen hoy. (pro), Está muy nerviosa.} \]

'Paul is accompanying Laura to see the exam results today. (pro = Laura) is very nervous (fem.).'

In (5), the third person plural ending of odian is enough to license a null subject co-referent with los vecinos, as it is clear from the verbal morphology that Juan cannot be the subject of the second sentence. As for (6), the feminine ending -a on the adjective nerviosa indicates that the subject of the second sentence is female, thereby eliminating Pablo as a possible subject. The presence of a singular feminine NP in the first sentence, Laura, in combination with the adjectival morphology, permits the existence of pro in this situation.

It is worth mentioning that the equivalent sentences in Puerto Rican Spanish are ungrammatical, and that other Romance varieties will allow the equivalent of one of the two sentences, but not both. This is a non-trivial observation for two reasons. First, it shows that even different varieties of what are commonly considered to be the same "language" can show significant structural differences. Second, it gives us insight into both the wide variation that exists within Romance, and also highlights how, within any one variety, different methods that favour certain potential subjects can produce different results in grammaticality when pro is used for co-referencing.

It is a principal motivation of this current work to identify how each particular Romance variety deals with the numerous ways of indicating and favouring subjecthood, and also to understand how the different Romance varieties compare with each other in their treatment of these factors. What we have seen here in Section 2.2.2 is just a hint of what awaits us ahead. We explore these issues in much greater depth in the following chapters.
2.2.3 The grammatical relations hierarchy

We return now to the issue of grammatical relations, and seek to investigate how they differ from each other. Greenberg (1966a: 37-8; cited in Croft 1990: 92) was one of the first to identify a linear ordering among the grammatical relations in his work on markedness patterns and typological categories, citing that subject ranked higher than direct object, which ranked higher than oblique4 with respect to certain syntactic and conceptual criteria.

What is most relevant to our study of the connection between grammatical relations and the licensing of the null subject is how different areas of syntax interact with the grammatical entities subject, object, and oblique, and so we analyse in this section the extent to which the various grammatical relations dictate what is syntactically permissible in the world’s languages.

There is significant evidence that the precise ranking of grammatical relations is motivated by how syntax treats these entities cross-linguistically. One such example is word order. Greenberg’s first universal proclaims that “in declarative sentences with nominal subject and object, the dominant word order is almost always one in which the subject precedes the object” (1966b: 77, cited in Croft 1990: 107). Exceptions do exist, as is usually the case whenever we talk about language; Greenberg himself does mention that the VOS and OVS word orders are attested5, and the order of indirect objects in relation to direct objects is not always consistent within the same language, such as with the English dative shift (Croft 1990: 107).

Another instance of the connection between the grammatical relations hierarchy and syntax can be found in a second hierarchy, the NP accessibility hierarchy. Keenan and Comrie (1977) observed the variation in the world’s languages with regard to restrictions on NP relativisation (Croft 1990: 108). English is an example of a language that is quite generous with regard to which NPs are able to be relativised: subjects, direct objects, indirect objects, and obliques are all

---

4 Just as there is no consensus regarding the definition of “oblique”, there is no wide agreement either as to how to treat the concept “indirect object”, or whether it is worthy of recognition as a separate entity at all. It is missing in Greenberg’s original hierarchy although, as we shall see later on in this section, it is included in other related hierarchies. For a detailed discussion of the merits of a discrete classification for “indirect object”, see Comrie (1989: 177) as a starting point.

5 Croft does offer that many of the languages with these word orders have ergative systems, allowing for the possibility that word order is best explained not with the grammatical relations we are discussing here, but instead by the ordering of absolutive/ergative (1990: 107). We choose not to dissect that debate here.
accessible to relativisation. On the other extreme is Toba Batak, an Austronesian language spoken in Sumatra, which permits only subjects to be relativised. Languages that are intermediate on this range also exist: Persian will allow subjects and direct objects to be relativised, and Tamil makes subjects, direct objects, and indirect objects available to relativisation. While it should not surprise us that not every single language obeys this ranking, the general data are consistent with the grammatical relations hierarchy:

Subject > direct object > indirect object > oblique

In other words, languages which allow those on the right side of the hierarchy to be relativised will also allow all the entities to its left to do the same. Conversely, if a language does not license the relativisation of a certain grammatical relation, it will also not permit anything to the right of it on the hierarchy to be relativised.

A third example of the syntactic relevance of the ordering of grammatical relations is the treatment of causative structures in Turkish. Comrie explains that Turkish is generally representative of other languages in this regard (1989: 175). In a Turkish causative sentence, the causer occupies the subject position, and so the causee appears as a direct object. When the non-causative verb in the construction is transitive and thus carries a direct object, the causee cannot appear as the direct object and instead must be manifested as an indirect object. To take this one step further, if the non-causative verb has its own indirect object, then the causee cannot occupy this position either, and must be expressed as an oblique object with the postposition tarafindan, equivalent to English “by”. What is apparent is that the causee is encoded in the leftmost position in the grammatical relations hierarchy not already occupied. The causee normally fills the subject position in its non-causative sentence, and in the causative equivalents, it will attempt to fill the direct object spot if it is available, and it not, then the indirect object, with the oblique adjunct its last resort if nothing else is available.

Now that we are satisfied that there is some precedent for the grammatical relations hierarchy to have a definite impact on a language’s syntactic system, we investigate whether this hierarchy is directly related to the licensing of the null subject.
2.2.4 Testing the factors

Here we test the various phenomena, grammatical relations and its related hierarchy, thematic roles, and switch reference, that we explored in depth in the previous sections. The reasoning behind our decision to test these factors is as follows. We know that subjects, direct objects, indirect objects, and obliques may show some differences with respect to certain syntactic constructions. We discussed how thematic roles are not strictly congruent to grammatical relations, but admit that NPs with particular thematic roles may be more likely to occupy certain syntactic positions in the sentence than others. Finally, we noted the Romance technique of dealing with switch references and how this is tied intimately with the possibility of having a null subject.

We consider all four grammatical relations and the five thematic roles mentioned in Section 2.2.1. This gives us twenty different combinations of grammatical relations with thematic roles. To induce instances of switch reference in the fifteen applicable cases, the first sentence will have a noun phrase with the appropriate combination of grammatical relation and thematic role, and this noun phrase will serve as the subject of the second sentence. Where possible, the thematic role of the subject of the second sentence was maintained as AGENT, both for the purposes of consistency but also because the thematic role AGENT is that which is most closely associated with the subject position of the sentence, so this may be the most telling in our investigation of the null subject. An attempt was made to keep the second sentences within each thematic role the same wherever it did not infringe upon comprehension; alternate sentences were used only if it was felt that maintaining the same sentence would only cause severe interference in naturalness without any increased benefit, thus strongly affecting grammatical judgements.

The English versions of the twenty sentence sequences are listed here below:

---

6 Remember that in situations where the subject is the grammatical relation being tested, this will be a case of same reference and not switch reference, since the subject of both sentences will be identical. As we are testing this for each of the five thematic roles, that leaves us with $20 - 5 = 15$ cases of switch reference.
AGENT

Subject: Peter read the poems at the funeral. (He) spoke very eloquently.

Dir Obj: They made Peter speak at the funeral. (He) spoke very eloquently.

Indir Obj: They made Peter read the poems at the funeral. (He) spoke very eloquently.

Oblique: The poems were read at the funeral by Peter. (He) spoke very eloquently.

THEME

Subject: Mike was arrested. (He) tried to escape but without success.

Dir Obj: They restrained Mike. (He) tried to escape but without success.

Indir Obj: They resisted Mike. (He) tried to advance but without success.

Oblique: They looked after Mike. (He) had eaten an omelette the day before and had fallen ill.

EXPERIENCER

Subject: Mark is afraid of dogs. (He) cries if he sees them.

Dir Obj: Dogs frighten Mark. (He) cries if he sees them.

Indir Obj: Dogs are disagreeable to Mark. (He) cries if he sees them.

Oblique: For Mark, dogs should be kept tied up. (He) cries if he sees them.

BENEFACTIVE

Subject: John was helped thanks to several donations. (He) can now pay for the operation.

Dir Obj: They helped John with much financial assistance. (He) can now pay for the operation.

Indir Obj: They applied the rules on disability payments to John. (He) can now pay for the operation.

Oblique: They collected a lot of money for John. (He) can now pay for the operation.
Subject: Paul received many cards from his friends. (He) thanked everyone for their kind thoughts.

Dir Obj: They spoiled Paul with affection. (He) thanked everyone for their kindness.

Indir Obj: They gave many gifts to Paul. (He) thanked everyone for their generosity.

Oblique: They prayed on behalf of Paul. (He) thanked everyone for their concern.

Appendix A lists the corresponding sentences in the six Romance varieties tested in this experiment: Brazilian Portuguese, Castilian Spanish, Catalan, European Portuguese, Italian, and Puerto Rican Spanish.

An examination of the relative acceptability of these sentences reveals that the licensing of pro is not determined by the specific grammatical relation or thematic role under consideration, but is based instead upon whether we are witnessing an instance of same reference or switch reference, and the language-specific rules regarding null subject usage in each case. Let us examine, as an example, the Puerto Rican Spanish items from Appendix A. We witness that the only grammatical sentences are 1, 5, 9, 13, and 17. We also note that these five are the only ones that display an instance of same reference; the other fifteen sentences are ungrammatical, and they are all examples of switch reference. This correlation is not coincidental.

We observe that, strictly speaking, the grammatical relation under consideration does not seem to have a direct influence on grammaticality here. While there is a divide between sentences with subjects as antecedents and those with non-subjects, there is no differentiation in grammaticality of the sentences among those with direct object, indirect object, and oblique antecedents. This would suggest that the crucially relevant factor is not grammatical relation per se, but same and switch referencing.

A similar argument can be made for thematic role. We notice that it has no direct effect on grammaticality for this set of sentences. We see that the sentences with subject antecedents are grammatical, whereas those with non-subject antecedents are not. This occurs regardless of thematic role. It does not matter whether we are dealing with an agent, theme, experiencer, benefactive, or recipient; the result is the same.

We will discover in subsequent chapters that there is, indeed, a distinction to be made between same and switch reference, and our data there will be consistent with those presented here in arguing that, for the purposes of licensing the null subject, there is no recognisable difference at all, in those languages that tolerate pro in cases of switch reference, between direct object, indirect object, or oblique case. In other words, these three are the

---

7 This is especially pronounced when we delve deeper, later on in this work, into Puerto Rican Spanish and Brazilian Portuguese.

8 This may be an argument in favour of the conservative definition of "oblique" as "any case other than the nominative (and vocative)" (cf. footnote 1). The results presented in this section do not make that claim indisputable by any means, but it does support the continuation of such a debate, and encourages further research on this particular topic.
syntactic positions where the NPs occupying those spots in the first sentence become the intended subjects of the second sentence, creating instances of switch reference.

It is safe to postulate that the permissibility of null subjects in instances of switch (and same) reference is not directly dependent upon the type of grammatical relation that the said subject was expressed as in the preceding sentence. This is also true of thematic roles, which show no influence on the licitness of pro; NPs of all five thematic roles in non-subject positions in the first sentence are equally able to serve as switch reference subjects in the second sentence. The interaction of the two is not relevant either. We therefore conclude that grammatical relations and thematic roles are not primary concerns in our attempt to understand better the licensing of the null subject.

2.3. **Animacy**

2.3.1 **The animacy hierarchy**

Silverstein, in his discussion on split ergativity, was one of the first to acknowledge the existence of a hierarchy involving animacy, proposing the following cascading binary system:

\[
\begin{align*}
+tu & -tu \\
+ego & -ego \\
+proper-proper & +human -human \\
+animate & -animate
\end{align*}
\]

(Silverstein 1976: 122)

Dixon expressed the same general concept but presented the hierarchy in a different fashion, choosing not to use Silverstein’s embedded system of two-way distinctions but instead offering an ordered ranking:

first-, second-person pronouns < third-person pronoun < proper names < human common noun < nonhuman animate common noun < inanimate common noun

(Dixon 1979:85, cited in Croft 1990: 112)

Croft (1990: 127) elaborated on Dixon’s framework by splitting the single hierarchy into four separate sub-hierarchies, each component being a relevant factor in the global animacy hierarchy. These four properties constituting the animacy hierarchy were described as such:
Although Dixon's and Croft's animacy hierarchies are quite similar, a number of differences between the two are worth pointing out. The first is that, in Dixon's hierarchy, there is an inherent assumption that everything at the level of "proper names" and below it (i.e., to the right of it on the hierarchy) can only have third person reference. While this is usually the case, it is not difficult to identify attestations where proper names and common nouns have first or second person reference. For instance, it is not clear how Dixon would classify in his framework the usage of Portuguese *o João*¹⁰ (as a second person proper name) or English "the present author" (first person human common noun) as they appear in the following sentences:

(7) *O João deseja mais alguma coisa para beber?*

'Do you (i.e., John) want anything else to drink?'

(8) The present author takes full responsibility for any errors in this work.

A second point of departure that directly relates to Croft's hierarchy (but which could also be extended to Dixon's proposal) is how it would treat entities that

---

¹ Note that the animacy sub-hierarchy intends to account for pure animacy (i.e., it measures animacy in the literal sense of the word). A tangent on the discussion in the main text is determining to what extent the animacy hierarchy actually has anything to do with animacy proper. For example, there is no intuitive reasoning as to why the first and second persons are anymore animate, in the literal sense, than the third person. The same can be said about pronouns and proper names that refer to the exact same entity: does the use of 'he' to refer to a man named 'John' make him anymore alive? It is beyond the scope of this work to examine the finer details of this debate, but it will suffice to say here that, where its precise meaning is crucial or unclear, explicit mention will be made as to whether the intended interpretation of "animacy" is its literal sense (such as the third sub-hierarchy presented above) or its broader sense (as in the term "animacy hierarchy" to refer to these typological frameworks).

¹⁰ The usage of the addressee's name in Portuguese with second person reference is an intermediate case on the politeness scale, somewhere between informal *tu* and the respectful *o senhor*. Such employment of a discourse partner's proper name as an alternative to a second person pronoun is not unique to Portuguese; this occurs in a number of Asian languages as well, such as Japanese and Thai, among many others.
are relatively low on one sub-hierarchy but relatively high on another. It is apparent
then that there is no method of establishing absolute rank; it must be admitted as well
that even relative rank is not absolutely predictable or indisputable. For example, it is
open to debate how an inanimate pronoun (e.g., ‘it’) would rank in comparison to a
human common noun (e.g., ‘the boy’), as the former is positioned higher on the NP
type sub-hierarchy, but the latter is superior on the pure animacy sub-hierarchy.

These two queries aside, it is accepted that a general ordering of animacy as
defined by Silverstein, Dixon, and Croft seems justifiable and reasonable. The next
step is to investigate the utility of establishing the existence of such a hierarchy. We
must identify whether this animacy hierarchy is purely descriptive, or whether this
ranking does indeed go beyond mere observations, and is meaningfully relevant to a
better understanding of the precise workings within a language.

2.3.2 The role of animacy in syntax

We saw in Section 2.2.3 that the ordering expressed in the grammatical
relations hierarchy does have a significant impact on the syntactic constructions
permissible in certain languages. Data were presented from a range of unrelated
languages to show that certain syntactic issues such as word order, the ability to
relativise, and causative formation, may be accounted for by the ranking established
in the grammatical relations hierarchy. We attempt a similar task in this section by
demonstrating that the animacy hierarchy has tangible ramifications on languages’
syntactic operations.

Case markings in Punjabi will provide some justification that the rankings
articulated in the animacy hierarchy are not wholly artificial. In this Indic language,
pronominal direct objects, animate nouns, and definite inanimate nouns require the
postposition nū, while it is not used with indefinite inanimate nouns. Not only is this
consistent with the proposal that those entities higher on the hierarchy are more likely
to be case marked, but it also lends credible support to the inclusion of definiteness as
an integral part of the animacy hierarchy (Croft 1990: 127).

In Mixe, a Mexican language belonging to the Mixe-Zoque linguistic family,
there are two different sets of verbal morphemes in use, depending on whether the
subject NP or the object NP is ranked higher on the animacy hierarchy (Foley and
Van Valin 1985: 288). The usage of one particular set of verbal affixes indicates that the subject is positioned higher on the animacy scale; if the object outranks the subject by the same criterion, a different set of morphemes is employed. Therefore, the verb will be different, depending on whether the sentence is “Peter hit the animal” or “The animal hit Peter”; in the former, the “subject higher” morpheme -t is used, whereas for the latter, the “object higher” morpheme y-...y (Foley and Van Valin 1985: 289).

Animacy does not have to affect the morphosyntactic system of a language; it may alter the choice of voice instead. Such is the case in Southern Tiwa, a Native American language spoken in New Mexico. In transitive constructions, when AGENT is first or second person, and is therefore no lower on the animacy hierarchy than THEME, the active voice is required. If THEME is first or second person, and therefore at least as high on the animacy scale AGENT, THEME is promoted to subject by means of passivisation, and so the passive voice is compulsory; the option to use the active voice with AGENT as subject is no longer available (Comrie 1989: 192-3). Note that what we have just seen in Southern Tiwa also incorporates thematic roles, suggesting that they are not just neutral observers in the licensing of syntactic constructions, but are indeed active participants.

Finally, we need not go any further than the language family we are concentrating on to find instances of the interrelation between the animacy hierarchy and syntax. Spanish, among others, marks a distinction between an animate and an inanimate direct object, in the literal sense, by introducing the former with the prepositional a, the so-called prepositional accusative. While it may appear at this point that we are back on the brink of the previously mentioned debate between the animacy hierarchy and literal animacy, it is worth noting here that the same construction applies in distinguishing between animate referential, and animate non-referential direct objects. This is illustrated by the following pair of sentences:

(9)  *El director busca a un empleado.*

(10)  *El director busca un empleado.*

‘The manager is looking for a clerk.’ (Comrie 1989: 134)
The two sentences are not strictly synonymous; in (9), there is an implication that a specific individual is being sought, whereas in (10), the manager is not looking for a particular clerk; any such employee will do.

The data presented in this section should be ample evidence to justify our claim that the animacy hierarchy does have some direct interaction with the syntactic operations of languages. We return to the relationship between the animacy hierarchy and syntax in Section 2.3.4 when we investigate whether this linear ranking has any effect on the licensing of the null subject. Before we do that, however, we must take a brief detour and discuss why we would even consider such a possibility in the first place.

2.3.3 Salience and topicality

It is legitimate to question why we should expect any sort of correlation between animacy (either in the literal sense or as proposed in the hierarchy bearing its name) and subjecthood. One factor that may influence the likelihood of an NP serving as the subject of the sentence is saliency. Foley and Van Valin assert that, cross-linguistically, those who participate in the speech act, the speaker and the addressee, are more salient than those who are absent, in other words, the grammatical third person\(^{11}\) (1985: 288). Furthermore, within third person noun phrases, humans are more salient than non-human animates, which in turn are more salient than inanimates. This hierarchy of inherent salience therefore ends up mirroring quite closely Dixon's animacy hierarchy that we examined in Section 2.3.1\(^{12}\). Foley and Van Valin then propose that "NPs higher on the inherent salience hierarchy tend to occupy more prominent syntactic positions than NPs lower on it" (1985: 288). By the term "more prominent syntactic positions", we can reasonably gather that what Foley and Van Valin intended was, among other things, the subject position of a clause. Thus, if entities that are higher on the animacy/inherent saliency hierarchy are more

---

\(^{11}\) However, here is no consistency cross-linguistically with respect to how the first and second persons are ranked. Foley and Van Valin (1985: 288) note that in Algonquian, the addressee is more salient than the speaker, whereas the opposite is true in Bantu. Dyirbal gives neither any preference in saliency.

\(^{12}\) Comrie (1989: 199) does warn us about the potential circularity of defining saliency and of relating it to animacy. It is not a primitive in itself, but instead is treated as the interaction of a number of factors, animacy being one of them. He mentions the same caution for topic-worthiness as well.
likely to occupy these prominent syntactic positions, then it is possible that they may be more easily accessible and/or identifiable in the discourse, and could then license *pro* more often than NPs lower on the scale.

Song (2001: 170) takes a similar approach but frames his argument in terms of topicality. The speech act participants are the most topical of the NPs, and there is a hierarchy of topicality that can be established: humans are more topical than animates, which in turn are more topical than inanimates. Wierzbicka claims that those NPs higher on the topicality hierarchy are “more interesting to talk about” than NPs that are less topical (1981: 67, cited in Song 2001: 170). It is understandable to hypothesise then that those entities that are more interesting to talk about tend to be focused more in communication and are likely to be more central to the discourse at hand. One of the possible results of this continued presence in the dialogue then is that these NPs should appear more frequently as the subject of a clause than other, less topical NPs. This tendency of more topical NPs to occupy the subject position might mean that they, too, are more readily accessed or identified in the discourse, and might permit a null subject in its place more often than an NP lower on the topicality hierarchy.

Finally, we note quickly that Silverstein’s cascading binary system, detailed in Section 2.3.1, is occasionally referred to as the Agency Hierarchy (Song 2001: 167), suggesting that the entities higher on his hierarchy are more agentive, and therefore more likely to act as subjects, by the reasoning we discussed in our exposition of thematic roles in Section 2.2.1. While we have seen in Section 2.2.4 that thematic roles have no direct bearing on the licensing of *pro*, the generally accepted assumption that animacy, agency, and subjecthood are all intimately interconnected is another motivation for our investigation of the animacy hierarchy and null subjects.

This section has attempted to demonstrate that there is value in examining whether the animacy hierarchy plays any role in the licensing of *pro*, as several links have been claimed between this ranking scheme and the likelihood of serving as a subject, these arguments being based on criteria such as salience and topicality. We now proceed to testing these effects.
2.3.4 Testing the factors

We return now to Croft's animacy framework. His hierarchy can be broken down into four distinct sub-hierarchies: person, NP type, (literal) animacy, and definiteness. For the sake of simplicity, we choose to set definiteness aside and focus our attention solely on the interaction of the other three factors. Each of these three properties has three different values\(^{13}\), so we end up with 27 different combinations, as outlined below:

possible values for “Person”: first, second, third
possible values for “NP type”: pronoun, proper name, common noun
possible values for “Animacy”: human, (nonhuman) animate, inanimate

1. first person, pronoun, human
2. first person, pronoun, animate
3. first person, pronoun, inanimate
4. first person, proper name, human
5. first person, proper name, animate
6. first person, proper name, inanimate
7. first person, common noun, human
8. first person, common noun, animate
9. first person, common noun, inanimate
10. second person, pronoun, human
11. second person, pronoun, animate
12. second person, pronoun, inanimate
13. second person, proper name, human
14. second person, proper name, animate
15. second person, proper name, inanimate
16. second person, common noun, human
17. second person, common noun, animate
18. second person, common noun, inanimate
19. third person, pronoun, human
20. third person, pronoun, animate
21. third person, pronoun, inanimate
22. third person, proper name, human
23. third person, proper name, animate

\(^{13}\) While the frameworks discussed in this work treat the first and second person as being equivalent, we choose here to examine them separately, each in its own right, to determine whether the commonly occurring clustering of speech act participants together in the literature can be supported by our investigation of null subject licensing.
24. third person, proper name, inanimate
25. third person, common noun, human
26. third person, common noun, animate
27. third person, common noun, inanimate

Of these 27 combinations, twelve of them are infeasible. We omit 2, 3, 5, 6, 8, and 9 above, as these are combinations of first person and non-human literal animacy. 11, 12, 14, 15, 17, and 18 are disregarded as they are the second person equivalents of the first set. In light of this anthropomorphic fallacy, the tendency to assign human attributes to non-human entities, it is unrealistic to consider a non-human speech participant. While one could claim for the second person that the addressee could be a non-human animate (e.g., speaking to the family pet), or even an inanimate object (e.g., “Where are you hiding?” in reference to lost house keys), this would be a mighty conceptual stretch, and its inclusion would not significantly add any insight to what we are attempting to discover. We feel justified then in leaving aside these twelve combinations.

That leaves us with fifteen combinations, as summarised below. The numbers in parentheses refer to the numbered combinations above:

**Table 1 – first person**

<table>
<thead>
<tr>
<th>human</th>
<th>non-human animate</th>
<th>inanimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pronoun</td>
<td>me (1)</td>
<td>X</td>
</tr>
<tr>
<td>Proper name</td>
<td>John Smith (4)</td>
<td>X</td>
</tr>
<tr>
<td>Common noun</td>
<td>the present author (7)</td>
<td>X</td>
</tr>
</tbody>
</table>

**Table 2 – second person**

<table>
<thead>
<tr>
<th>human</th>
<th>non-human animate</th>
<th>inanimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pronoun</td>
<td>you (10)</td>
<td>X</td>
</tr>
<tr>
<td>Proper name</td>
<td>John Smith(^\text{14}) (13)</td>
<td>X</td>
</tr>
<tr>
<td>Common noun</td>
<td>the doctor(^\text{15}) (16)</td>
<td>X</td>
</tr>
</tbody>
</table>

\(^{14}\) See footnote 10.
We present below the English versions of the sample sentences containing noun phrases with each of these combinations. In each item, the noun phrase to be tested is in an oblique construction in the first sentence, and is meant to be the subject of the second sentence\textsuperscript{16}. The subjects of the second sentence were given the thematic role AGENT. The sentence numbers correspond to the numbering in the combination list and the tables above.

1. All responsibility rests with me. (I) must accept the consequences.
4. All responsibility rests with John Smith. (I) must accept the consequences.
7. All responsibility rests with the present author. (I) must accept the consequences.
10. All responsibility rests with you (You) must accept the consequences.
13. All responsibility rests with John Smith. (You) must accept the consequences.
16. All responsibility rests with the doctor. (You) must accept the consequences.
19. All responsibility rests with him. (He) must accept the consequences.
20. The baby is crying because of it. (It) scared him.
21. The baby is crying because of it. (It) scared him.
22. All responsibility rests with John Smith. (He) must accept the consequences.

\textsuperscript{15} The usage of a common, referential noun in Portuguese is parallel to the same with proper names.

\textsuperscript{16} We claim here, based on what we saw in Section 2.2.4, that the grammatical relation of the NP in the first sentence is inconsequential. We select the oblique construction throughout for the sake of consistency.
23. The baby is crying because of Rex. (It) scared him.

24. The baby is crying because of the Picasso. (It) scared him.

25. All responsibility rests with the boy. (He) must accept the consequences.

26. The baby is crying because of the dog. (It) scared him.

27. The baby is crying because of the painting. (It) scared him.

Appendix B shows the corresponding sentences in Brazilian Portuguese, Castilian Spanish, Catalan, European Portuguese, Italian, and Puerto Rican Spanish. We see that whether the null subject is permitted is independent of NP-type and animacy, but that grammatical person does show some effect. Examining our Puerto Rican Spanish items in Appendix B as an example, we observe that sentences 19 through 27 are all ungrammatical. Note that these nine sentences, all with third person subjects, span the possible combinations of NP-type (pronoun, proper name, and common noun) with values for animacy (human, non-human animate, and inanimate). So the NP-type and animacy of the subjects in question have no bearing on determining the grammaticality of the sentence.

Now we observe that sentences 1, 4, and 7, with first person subjects, and sentence 10, with a second person subject, are grammatical. The sentences corresponding to these with third person subjects (namely, sentences 19, 22, and 25) were judged to be ungrammatical. This hints to the positive role that grammatical person may play in licensing null subjects, and will be examined in greater depth in subsequent chapters.

We conclude then that neither the animacy hierarchy itself, nor any pair-wise combinations of the four properties that constitute it, is directly relevant to the licensing of pro. The licitness of the null subject appears to be independent of the NP type or literal animacy of the antecedent of pro.

This result should not be entirely surprising. As we have mentioned before, there is nothing to indicate that the speech participants are necessarily any more animate literally than a human third person, so any differences in the licensing of pro arising from the effect of grammatical person is not correlated to animacy, but is attributable to other factors. Additionally, a pronoun, a proper name, or a common noun referring to the exact same entity should not show any animacy differences, because all three are being applied to one and the same. While the most interesting discovery from our investigation is that NPs referring to humans, non-human animates, and inanimates behave the same way as well with respect to the licensing of the null subject, it should not come as an unexpected shock then that the animacy hierarchy does not have any direct effect on the licensing of pro, considering that most of the parts that make up that framework individually do not either.

We move away from typology, hierarchies, and synchronic treatments of language and focus now on historical linguistics. In the next section, we examine the development of subject pronoun usage in the history of the French language, and investigate how it may be relevant to our understanding of null subjects in contemporary languages.
2.4 Subject pronoun usage in the history of the French language

2.4.1 Word order and null subjects in Old and Middle French

Old French displayed a number of syntactic properties that are not present in its modern-day counterpart. One of these is the licensing of simple inversion. Word order was very flexible in old French, to the extent that:

“any of the six permutations of subject, verb and object were possible, though some were more common than others. [...] Inversion of verb and subject, however, was particularly frequent, being normal whenever an adverb or adverbial expression, or a grammatical object, opened the sentence or the clause.”

(Rickard 1989: 54).

So a sentence with a structure such as that in (11a) below was grammatically acceptable, whereas its strict Modern French equivalent (11b) is not:

(11a) Sont les bestes de l’air si abandonees que vos les doiez ocirre sans reson? (La queste del saint Graal 97, 32, cited in Bonnard and Régnier 1989: 203)

(11b) *Sont les bêtes de l’air si rejetées que vous deviez les abattre sans raison?

‘Are the beasts of the air so rejected that you had to kill them without a reason?’

What is required to render a sentence such as (11a) grammatical in the modern language is to do one of two things. The first option is to use complex inversion; that is, to insert a pronoun co-referential to the subject of the sentence, and then to invert this pronoun and the verb after the R-expression, as in (11c) below. The other choice is demonstrated in (11d), the addition of the non-emphatic interrogative marker est-ce que before the subject of the sentence, a construction dating from the 15th century.
(Foulet 1921, cited in Roberts 1993: 143), that leaves the canonical word order of the subject followed by the verb intact.

(11c) *Les bêtes de l’air sont-elles si rejetées que vous deviez les abattre sans raison?*

(11d) *Est-ce que les bêtes de l’air sont-elles si rejetées que vous deviez les abattre sans raison?*

Another characteristic of Old French not observable in Modern French is the V2 nature of its clauses. Sneyders de Vogel (1927: 382) noted that “*[e]n vieux français – on n’a qu’à lire quelques pages de Villehardouin – il y a une forte tendance à mettre le verbe à la seconde place de la phrase*”18. This is clearly illustrated in (12a) below:

(12a) *Quatre saietes ot li bers au costé* (Charroi de Nîmes, 1.20, cited in Roberts 1993: 85)

Four boats of war had the baron at his side

‘The baron had four boats of war at his side.’

With the exception of a few relic constructions (e.g., *Peut-être me suis-je trompé*; *Veut-elle venir*?), this V2 rule does not apply any longer in French, so that the strict equivalent in the contemporary, everyday language is ungrammatical:

(12b) *Quatre bateaux de guerre avait le baron à son côté.*

(12c) *Le baron avait quatre bateaux de guerre à son côté.*

---

18 “*In old French – one only has to read a few pages of Villehardouin – there is a strong tendency to put the verb in the second position of the sentence.*”
A third feature that separates Old French from Modern French is the licit omission of subject pronouns in the former. Citing textual evidence from the early 12th century *La Chanson de Roland*, Ayres-Bennett asserts that “subject pronouns could also still not be expressed […] even if there was a new subject introduced in the clause” (1996: 66). This point can be illustrated with an example from *Perceval* in (13a) below, where the subject of the first verb is covert. It is well known that French is peculiar among its Romance relatives in that subject pronouns must always be expressed (except in imperatives), as shown in (13b):

(13a) Sire, ne sai se je suis pres

(Le Conte du Graal 1576, cited in Bonnard and Régnier 1989:46)

(13b) Sire, *(je) ne sais si je suis près

‘Sir, I don’t know if I’m close’

During the middle French period19, all three of these syntactic characteristics in the language underwent changes. The first of these mentioned above, simple inversion, was no longer required. Although it was still permitted until the early 16th century, it was gradually replaced by complex inversion and the addition of *est-ce que* (Roberts 1993: 192). Rickard notes that:

“[An] important development in word-order is that inversion is no longer automatic after a preceding adverb or adverbial phrase. It is still quite frequently found in the fifteenth century, but the direct word-order is also common, and particularly so when the subject is a pronoun.”

(Rickard 1989: 72)

The second property, the V2 nature of clauses, was also in a state of transition. In Robert’s quantitative examination of the frequency of V>2 orders, where at least

---

19 There is no universal agreement on when the middle French period begins and ends. Just to mention a couple of the possible ranges that have been postulated, Rickard (1989: 61) claims that there is consensus that it stretches from the first half of the 14th century until the first half of the 17th century. On the other hand, Ayres-Bennett (1996: 98) allows that the middle French period may have begun anytime between the middle of the 13th century to the end of the 14th century, and may have ended sometime between the last quarter of the 15th century and the first third of the 17th century. She elects in her work to delimit this period as lasting from the beginning of the 14th century until the end of the 15th century.
two maximal constituents come before the inflected verb, he reveals that $V>2$ order occurs two to three times as often in the middle French texts studied than in Old French texts (1993: 148). It was optional in the language until the early 16th century (Roberts 1993: 153), the period at which Adams, among others, claims that the V2 word order disappeared from the language (1987: 27).

The third aspect of Old French setting it apart from Modern French, the permissibility of null subjects, is also in the process of being transformed. It starts to be used regularly enough in the 14th and 15th centuries (Sneyders de Vogel 1927: 44), but does not become obligatory until a couple of centuries later. Examples from texts such as the mid-15th century Cent nouvelles nouvelles document this rise in usage. Ayres-Bennett refers to it when she comments that it:

"[...] illustrates well the increased use of subject personal pronouns during Middle French; while we are not yet at the stage of having a subject pronoun with every verb, as was required by seventeenth-century grammarians, they appear much more regularly than in Old French texts, and feature, for instance, in inverted structures."

(Ayres-Bennett 1996: 106)

So by the end of the Middle French era, the language had experienced a number of transformations that made it quite different from how it was at the beginning of that period. Alternate strategies were used to replace the now ungrammatical simple inversion, V2 word order was no longer compulsory, and subject pronouns were increasingly employed until they were finally made to be obligatory. The relevance of these three changes is discussed below.

2.4.2 Switching of a parameter

The three syntactic features discussed in Section 2.4.1, namely, simple inversion, V2 word order, and null subjects, while seemingly distinct properties, are not in fact fully independent of each other. Foulet was one of the first to identify their interrelation when he observed a link between inversion and null subjects in the early 20th century, stating that "l'inversion du sujet entraîne facilement dans le cas du
pronom personnel l’omission du sujet” (1919: 313). Price offers that “the tendency to place the verb in the second position and the possibility of dispensing with the subject pronoun, must be borne in mind if the use of unstressed subject pronouns in [Old French] is to be properly understood” (1998: 146). Adams provides evidence in support of this, suggesting that the permissibility of null subjects is directly correlated with the existence of V2 order (1987: 25-27), and so the elimination of V2 order from the French language by the 16th century entails a corresponding increase in overt subjects.

The reason for the connection between simple inversion, V2 word order, and null subjects has a syntactic justification. The common thread tying the three together is that in each of these phenomena, “the inflected verb appears in C and licenses or Case-marks the subject from that position” (Roberts 1993: 86). These syntactic changes, then, can be explained by a single unified argument. Roberts proposes that the loss of simple inversion, V2 word order, and null subjects in French is accounted for by a change in a parameter setting, namely, that Αγρ used to be able to assign Nominative case under government, but that in the modern language this parameter is now set negatively so that this assignment is no longer possible (1993: 81).

A diachronic look into the situation will highlight the fact that ambiguity played a significant role in the changing of this parameter setting. For instance, in the case of inversion, Roberts presents examples from 16th and 17th century texts to illustrate that there were a number of sentences where it was ambiguous as to whether it was an instance of simple inversion or of free inversion, and there were several undisputable tokens of free inversion (1993: 191). Therefore, it was reasonable and intuitive for all of the ambiguous cases to be treated as free inversion, thus expunging simple inversion from the grammatical system.

As for word order, there was confusion about the status of SVO clauses. Marchello-Nizia calculates that the SV word order constituted anywhere from one-half to three-quarters of the cases in the Middle French texts consulted, whereas the corresponding fraction for the same in Old French was generally no higher than one-half (1979: 331). So the SVO word order had clearly become favoured, as it was now

20 “Inversion of the subject easily brings about the omission of the personal subject pronoun.”
21 Αγρ can assign Nominative under agreement as well, but Roberts mentions that this particular parameter doesn’t change during the historical development of French, remaining set positively still to the present day (1993: 88).
22 It should be noted that Roberts did find just one incontestable example of simple inversion in the 16th century texts he examined, and none in the 17th century, in d’Aubigné’s Lettres (1993: 190).
able to account for a majority of the constructions examined. The ambiguity that arose at this stage was whether the language could be characterised as an SVO V2 system or just a SVO one. Adams offers that it was the ambiguity of these increasingly frequent SVO clauses found in the language that caused it to be reanalysed as a simple SVO system (1987: 25-26). Matrix clauses were now understood to be AgrP and no longer CP, and while \( C^0 \) continued in late Middle French to carry the feature [+Agr], the eventual loss of that feature on \( C^0 \) brought about the disappearance of V2 in the early 16\(^{th} \) century (Roberts 1993: 187; 199).

The end result is that, since the presence of null subjects was dependent on the licensing of simple inversion and on a V2 system, the loss of these two properties entailed the end of covert subjects in the French language. The parameter determining whether \( \text{Agr}^0 \) could assign Nominative case under government switched from 'yes' in Old French to 'no' in Modern French, so that simple inversion was no longer permitted and the V2 order was not a requirement any longer in the language. These particular changes brought about a drastic reduction in the number of syntactic contexts in which null subjects were licensed (Roberts 1993: 204), so that \textit{pro} could no longer appear in positions where it used to be allowed previously.

2.4.3 Verbal morphology and assumptions

Sections 2.4.1 and 2.4.2 highlighted the developments that the French language experienced with respect to inversion and word order, and attempted to clarify how these syntactic changes, caused by a resetting of a certain parameter, directly affected the rules on the usage of subject pronouns. This section investigates the alleged role that French verbal morphology plays in the present-day requirement of overt subject usage, and how these two ideas are consistently, and erroneously, linked.

There is a common misconception that it is the weak phonological agreement of morphological endings in the French verbal paradigm that has sparked the use of subject pronouns. Consider a regular verb in the dominant -\textit{er} conjugation group. Its forms for the present indicative are as follows, with phonetic transcriptions of the verbs in brackets:
J'aime  [em]  I love
tu aimes  [em]  you (singular informal) love
il aime  [em]  he loves
nous aimons  [emō]  we love
vous aimez  [eme]  you (plural/polite) love
ils aiment  [em]  they love

All three singular persons, and the third person plural, are homophonous; historically, this is a result of the elimination of the Latin verbal terminations –s, -t, and –nt for the second person singular, third person singular, and third person plural verbal forms respectively. Foulet claims that these four persons have been phonetically identical since as early as the 12th century (1935/6: 292, cited in Roberts 1993: 126).

Some historical grammars attribute the usage of subject pronouns to this lack of phonetic distinction:

"Plus donc les désinences s'affaiblissent, plus la langue...a besoin d'autres moyens pour exprimer les personnes, et plus l'emploi du pronom personnel devient général...le pronom sujet s'introduit donc de plus en plus pour remplacer les terminaisons qui se perdent."23

(Sneyders de Vogel 1927: 42-3)

"En effaçant de nombreuses désinences verbales en moyen français, la phonétique a rendu le pronom indispensable, d'abord au singulier et à la 3e personne pluriel [...] puis, par analogie, aux 1re et 2e personnes pluriel [...]"24

(Dauzat 1930: 414-415)

What may come as more of a surprise, however, is that even some contemporary French reference grammars make such a claim:

23 "Therefore, the more the desinences get weakened, the more the language needs other means to express person, and the more the use of the personal pronoun becomes general [...] the subject pronoun is introduced, therefore, more and more in order to replace the terminations that are lost."

24 "By erasing the numerous verbal desinences in Middle French, phonetics has rendered the pronoun indispensable, first in the singular and the third person plural [...] then, by analogy, to the first and second persons plural [...]"

58
"The most important syntactical consequence of the equalization of verb endings in the oral code is that in modern French an explicit subject is required in all sentences other than imperatives."

(Hollerbach 1994: 155)

"Avec l'affaiblissement des désinences, on verra s'étendre, de plus en plus, le pronom personnel sujet...le pronom sujet est devenu indispensable pour marquer la personne."

(Perret 1999: 130)

It is important to clarify that we are not saying that the elimination of phonetic distinctions on these verbal endings is completely unrelated to the required usage of subject pronouns in the modern language. What needs to be emphasised is that this syncretism of the singular and the third person plural in the French verbal paradigm is not the actual cause of the overt subject requirement; the cause of obligatory subject pronouns is explained by the changes in inversion and word order in the historical development of the French language, as we saw above. Indeed, the merging of these four originally different inflectional forms may well have reinforced or supported the necessity of subject pronouns. Price summarises this possibility quite clearly:

"It is sometimes suggested that the growth of the use of the subject pronouns in French, a feature that differentiates French from other Romance languages, in which the pronoun is little used except for purposes of clarity or emphasis [...] is a consequence of the loss of the personal endings of the verb: when forms [...] came to be pronounced alike, the language had recourse to the subject pronouns to differentiate between the different persons on the verb. However, things were perhaps not as simple as this. [...] The loss of personal endings, though certainly not at the origin of the use of unemphatic [subject pronoun] [bold PSM], can only have encouraged an already well-established practice to have become an almost invariable one."

(Price 1998: 148-9)

We have attempted to show up to this point that the disappearance of formerly distinct verbal endings did not result in subject pronouns being used compulsorily, and have provided a succinct overview of the extent to which these two phenomena

25 "With the weakening of the desinences, one sees the use of the subject personal pronoun increasingly spreading [...] the subject pronoun has become indispensable in indicating person.”
are viewed to be formally linked, both in older literature as well as in more recent reference grammars. This provides an apt transition to the next subsection, where we look at a possibly comparable situation in modern Caribbean Spanish, and examine how far we can draw parallels between it and what we have observed in the history of French.

2.4.4 Modern Caribbean Spanish: déjá vu?

Suñer states that final consonants in Caribbean Spanish\(^{26}\) are weakened, at times to such an extent that they are unperceivable in unmonitored speech, so that the phonetic reduction or elimination of terminal \(-s\) causes the neutralisation of previously distinct second and third person singular forms (1986: 196). Consider that in the imperfect indicative, the conditional, the present subjunctive, and in both the \(-ra\) and the \(-se\) forms of the imperfect subjunctive, the first person singular is formally identical to the third person singular. Combine that with the observation that final \(-n\) may be weakened as well, causing at least near homophony with the third person plural\(^{27}\), and we are able to remark that, for those selected verbal paradigms, all forms of the singular and the third person plural may be indistinguishable to the relaxed ear. Note that these are the same four persons involved in the auditory syncretism we witness in most Modern French verbal paradigms.

Suñer then connects this erosion of terminal consonants to the ubiquity of subject pronouns in Modern Caribbean Spanish. She writes that, with regard to these weakened final consonants, “related to this […] issue, and probably prompted by it

\(^{26}\) Suñer does not specify exactly what part of the Caribbean this variety of Spanish is representative of. It is assumed by the present author that Suñer treats all varieties of Spanish originating from that region as being the same, although she does not explicitly say this herself anywhere in her work. In one short paragraph, there is a brief mention of attestations observed while she was in Caracas (Suñer 1986: 197), but it is not apparent whether all data from her entire work are based on Venezuelan Spanish, or just those that were cited in that paragraph. This brings us to another question, whether Venezuela is considered a part of the Caribbean or of South America; it is usually classified as belonging geographically to the latter (such as in Mur-Molina 1997: 16). Returning to the main point, it should be noted that there may be arguments in favour of treating the varieties of Spanish spoken in the Caribbean as different entities. See, for example, Lipski (1994: 233; 241; 335) for syntactic differences between Cuban, Dominican, and Puerto Rican Spanish, in particular with regard to the system and usage of subject pronouns. For the sake of consistency, in this work we stick with Suner's use of the term “Caribbean Spanish” throughout the section.

\(^{27}\) Nasalisation of the vowel preceding the dropped \(-n\) often occurs, so that sometimes a distinction between the third person plural form and the other forms can still be detected even if terminal consonants are omitted (Suñer 1986: 196).
[bold PSM], is the fact that speakers of this dialect use subject pronouns much more frequently than do speakers from non-Caribbean regions" (Suñer 1986: 196). Suñer goes on to mention that subject-verb inversion in interrogatives no longer occurs obligatorily, and that there is attested usage of “QU-word – ser – (lo) - que”, the Caribbean Spanish equivalent of the French qu’est-ce que, as an alternative interrogatory strategy to subject-verb inversion (1986: 196-7).

Suñer explicitly links non-inversion and the existence of this interrogative marker to each other, but she does not relate these two phenomena directly to the increased use of subject pronouns in Caribbean Spanish (1986: 197). It may indeed be surprising that Suñer does not mention any parallels between what is observed here and what we have discovered by looking at the history of the French language, especially considering the striking similarities between these two situations.

Two comments are in order here. The first is that we see again the suggestion, either direct or implied, that the loss of strong agreement within a verbal paradigm is the principal cause for the introduction of overt subject pronouns to disambiguate. There is a precedent for keeping strong agreement and subject pronoun usage as separate ideas independent of each other. For instance, Roberts contributes that “the agreement system of Middle French was fairly poor, but in essence neither significantly poorer nor significantly richer than that of Old French and at least early Modern French” (1993: 185), yet we see great diversity in the way subject pronouns are used in the history of the French language. Considerable effort should be made to remember that there may be reasons other than the apparently obvious verbal syncretism, such as word order, that explain the licensing of pro. It would be prudent when considering, for example, Lipski’s assertion that the retention of subject pronouns in Puerto Rican Spanish is caused at some level by “the erosion of final consonants which signal verbal morphology” (1994: 335), that non-inversion of subject and verb after QU-words in interrogatives, which Lipski does mention but does not link to subject pronoun usage, may be a relevant factor.

The second comment is to reflect on the timing of these various events. French remained a null-subject language until the 16th century, hundreds of years after the loss of auditory distinction of the four verbal endings in question (Roberts 1993: 126), so this would strongly favour the argument against directly linking verbal morphology and subject pronoun usage. Similarly, “null subjects disappear from this configuration essentially at the same time as simple inversion” (Roberts 1993: 216),
so it would be worth investigating whether these similar developments in Caribbean Spanish are contemporaneous with each other, as this could provide insight on the link between these syntactic changes.

What is crucial to keep in mind, then, is that weakened verbal morphology does not always correlate directly with the licensing of pro. The loss of discrimination among verbal terminations does not necessarily entail the introduction of subject pronouns to disambiguate what person is intended. The benefit of examining a situation such as the development of subject pronoun usage in the history of the French language is that it provides a sound lesson that may be applied to modern situations. We may look at how these changes came about in French as a historical precedent to help guide us towards a better, and more accurate, understanding of the parallel developments in Caribbean Spanish.

2.5 Summary

This chapter has been devoted to identifying factors that, at first glance, would appear to affect the presence of the null subject, but that actually do not. We first examined in Section 2.2 the supposed correlation between grammatical relations and thematic roles, and discovered that neither of these properties, nor the interaction of the two, had any direct influence on the licensing of pro. In that section, we also took a brief look at the phenomenon of switch reference, and investigated the grammatical relations hierarchy. In Section 2.3, we continued our scrutiny of typological hierarchies, discussing the various approaches to the animacy hierarchy, and how it related to salience, topicality, and therefore subjecthood. We also discovered that neither the animacy hierarchy, nor literal animacy itself, contributed to the licitness of the null subject. In both of these sections, we presented data from a number of Romance varieties illustrating our point. We followed this in Section 2.4 with a brief recapitulation of the history of subject pronoun usage in French, and accepted that its required use in the modern-day language was a result of changes in word order and interrogative strategies, and not by reduced verbal morphology, as is often claimed. We applied this lesson to the current situation in Caribbean Spanish, whose constructions show a path of development similar to what occurred in French.
Now that we have explored a handful of factors that are unrelated to the principal topic of this work, it is time to contemplate those properties that do indeed directly affect the Romance null subject. In the chapters that follow, we identify what these conditions are, examine how each individual Romance variety addresses them, and consider the range of different methods of treatment employed by the Romance languages.
Chapter 3: Literature review and methodology

3.1 Introduction

Before presenting the data collected for this dissertation and initialising an investigation of how this research supplements our current knowledge of the null subject and other related linguistic issues, we first summarise the previous accounts that have attempted a cross-linguistic explanation of the null subject phenomenon. The purpose of this is not only to clarify, but also to justify, the methodological motivations and intended goals of the present work. It is hoped that the reader, via this survey of literature, will acquire a better understanding of the specific issues we attempt to address and explain in the current work. Furthermore, by validating the approaches and reasoning taken in this dissertation, we claim that this can only strengthen the arguments and conclusions presented later on this work.

We first provide a brief synopsis of earlier approaches to understanding cross-linguistic variation in the null subject, highlighting in particular their lasting contributions and also their shortcomings. From there, we then continue by examining in much greater detail some more recent works that have directly influenced both the empirical structuring and the theoretical aims of this dissertation. This provides a useful basis from which to begin our discussion of the unique features of the current research.

It should also be pointed out that this brief review is not intended to be exhaustive; indeed, it cannot be. To attempt a thorough retrospective of even just a fraction of all the significant research on this topic would require more space than we have at our disposal. Rather, the intended task of the rapid recapitulation offered here is to focus briefly on the handful of contributions that have had a tangible influence, either by their theoretical approach to the topic or by the actual investigative techniques they use in their exploration of the null subject phenomenon, on the present research or on those works upon which this one is based.
3.2 Gilligan (1987)

3.2.1 Null subjects, subject inversion, and that-trace filter violations

Gilligan's work is often credited as one of the first to pursue a wide and systematic cross-linguistic study of the null subject. He achieves this in two separate surveys. In the first, he examines four properties relating directly or indirectly to the null subject, and by resorting to data collected from a selection of a hundred languages, a sampling proportionally mirroring the distribution of the world's language families as suggested in Ruhlen (1987), he attempts a typology of null subject usage. The four properties he considers, following proposals by, among others, Perlmutter (1971), Chomsky and Lasnik (1978), Kayne (1980), Rizzi (1982), and Safir (1985), are:

i) pro – whether the language permits a null thematic subject  
ii) EXE – whether the language permits a null non-thematic subject  
iii) SI – whether subject inversion is permitted  
iv) THAT – whether the language exhibits that-trace filter violations

(Gilligan 1987: 130)

One of two binary values is assigned for each of the four properties in each language, or if a certain property cannot be evaluated, the designation ND (not determinable) is given. Even a cursory appreciation of the cross-linguistic variation established by just these four factors can be gathered from the data provided for, to choose three unrelated languages, Italian, which has all four marked as 'yes'; Papiamentu, which has pro and SI marked as 'no' and the other two as yes; and Finnish, which has pro and EXE marked positively and the other two negatively (Gilligan 1987: 135).

From his vast table of results, Gilligan is able to compare these four properties pair-wise in order to investigate implications and correlations. One of the ideas he considers is the hypothesis that subject inversion and that-trace filter violations are correlated with the existence of null subjects, as proposed by Rizzi (1982) and Safir.
(1985), and discovers that their relationship is rather weak. The numbers below summarise this:

<table>
<thead>
<tr>
<th></th>
<th>[SI]</th>
<th></th>
<th>[THAT]</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>[pro]</td>
<td>yes</td>
<td>22</td>
<td>49</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>no</td>
<td>11</td>
<td>15</td>
<td>no</td>
</tr>
</tbody>
</table>

(Gilligan 1987: 144) (Gilligan 1987: 146)

The instances of counterexamples in the survey are striking. In the comparison between pro and SI, Gilligan uncovers eleven languages that permit subject inversion but not a thematic null subject, and 49 languages that allow the thematic null subject but not subject inversion. This is hardly a strong endorsement for the bundling of these two qualities. As for Gilligan’s investigation of null thematic subjects and that-trace violations, he identifies at least one language in each of the four theoretical combination patterns, but the overwhelming number of languages (89 of them) that are not analysable as a result of at least one ND setting makes this correlation inconclusive at best, and extremely unlikely at worst.

With regard to the other four pair-wise comparisons, the zeros in three of those four charts below do lead to informative implications:

<table>
<thead>
<tr>
<th></th>
<th>[EXE]</th>
<th></th>
<th>[THAT]</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>[pro]</td>
<td>yes</td>
<td>24</td>
<td>0</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>no</td>
<td>15</td>
<td>2</td>
<td>no</td>
</tr>
</tbody>
</table>

(Gilligan 1987: 136) (Gilligan 1987: 146)
The comparison between pro and EXE makes it reasonably credible that null thematic subjects imply null non-thematic subjects. While less convincingly substantiated on the basis of lower numbers, it is a plausible suggestion that subject inversion implies *that*-trace violations, which in turn implies null non-thematic subjects. Gilligan summarises Safir’s (1985) analysis of the correlation as follows:

\[
\begin{align*}
\text{Pro} & \rightarrow \text{EXE} \\
\text{SI} & \rightarrow \text{THAT} \\
\text{THAT} & \rightarrow \text{EXE} \\
\text{SI} & \rightarrow \text{EXE} \text{ (by transitivity of the two statements directly above)}
\end{align*}
\]

(Gilligan 1987: 146-7)

### 3.2.2 The correlation between null subjects and rich agreement

In Gilligan’s second survey, he investigates, in the same one hundred languages, the following eight entities:

i) Sthem – thematic subject of a finite clause  
ii) Sinf – subject of a non-finite clause  
iii) Simp – subject of an imperative  
iv) EXE – non-thematic (expletive) subject  
v) Dobj – direct object  
vi) Iobj – indirect object

---

1 But see Raposo and Uriagereka (1990: 513) and Ledgeway (2000: 77-8) for counterexamples in Galician and Neapolitan, respectively, pertaining in particular to pragmatic requirements on the overt usage of expletive subject pronouns.
vii) Poss – possessive pronoun
viii) PPobj – object of adposition

(Gilligan 1987: 188)

For each of these eight linguistic items, in all of the languages surveyed, one of four values is assigned, based on all combinations of two separate parameters, whether that entity can be represented by a null pronoun, and whether there is feature agreement between this pronoun and another local linguistic entity that leads to a pronominal interpretation:

i) $\emptyset + A$ - a null pronoun may appear here, and agreement is present
ii) $\emptyset - A$ - a null pronoun may appear here, and agreement is absent
iii) $*\emptyset + A$ - a lexical pronoun must appear here, and agreement is present
iv) $*\emptyset - A$ - a lexical pronoun must appear here, and agreement is absent

(Gilligan 1987: 187)

Of the eight items examined in this second survey, we are most interested in Sthem, as this will reveal the extent to which a null representation of the thematic subject is correlated with the presence of morphological agreement between these pronouns and the verb. The result table is presented below:\(^2\):

\[
\begin{array}{ccc}
\text{[agreement]} & \text{yes} & \text{no} \\
\text{[Sthem]} & \text{null} & 76 & 17 \\
& *\text{null} & 2 & 9 \\
\end{array}
\]

(Gilligan 1987: 196)

---

\(^2\) These numbers add up to 104, not 100, because four languages have been classified in two different categories.
We see conclusively that the relationship between null thematic subjects and verbal agreement is not "biconditionally correlated" (Gilligan 1987: 160). While there are 76 languages supporting null subjects with verbal agreement (the $\emptyset + A$ designation above) and 9 languages lacking both null subjects and verbal agreement (*$\emptyset - A$), there are nineteen languages to serve as counterexamples, seventeen with null subjects but lacking verbal agreement ($\emptyset - A$) and two without any null subjects but displaying verbal agreement (*$\emptyset + A$).

Gilligan is thus able to evaluate systematically what was at the time a generally held belief, albeit one that was recognisably in need of much greater refinement, that a language with rich desinences in its verbal paradigm would be able to license null subjects, and that, likewise, poor agreement morphology would prohibit them, reflecting various proposals by Perlmutter (1971), Taraldsen (1978), Pesetsky (1982), Huang (1984), and Rizzi (1986). While postulating initially that this hypothesis cannot be sustained, mentioning as a token example the situation in Bavarian German, where null thematic subjects are permitted even though that language shares the exact same set of agreement morphemes as standard German, a variety that does not allow null thematic subjects (Gilligan 1987: 170), Gilligan provides substantial and conclusive evidence to further his opposition that null subjects and subject agreement morphology are obligatorily contingent upon each other.

3.2.3 A pro-drop typology

Based significantly on the results above, Gilligan posits four\(^3\) classifications of languages based on how accommodating they are to null subjects (1987: 398ff). Languages may belong to the core group of NSLs (Italian, Spanish, and Thai are included in this category), restricted NSLs (e.g., Classical Arabic and West Flemish), EXE-NSLs

---

\(^3\) Gilligan speculates on the possibility of a fifth category, which would include Dutch, based on the position of the thematic subject in that language (1987: 414). It is not clear, however, whether this hypothesis is offered because Dutch truly deserves a unique classification, or whether Gilligan is just unsure if it should be classified in EXE-NSL or non-NSL, as Dutch displays characteristics straddling these two groups.
(e.g., standard German and Icelandic), and non-NSLs (e.g., English, French, and Swedish).

Gilligan’s core NSLs are those where “non-thematic subjects are always null and thematic subjects always [both bold PSM] have the option of being null” (1987: 399). Restricted NSLs and EXE-NSLs share the property that null non-thematic subjects are obligatory in certain syntactic contexts, while prohibited in other environments. The difference between the two is that the former group permits null thematic subjects, whereas the latter does not license them. Finally, the non-NSLs are languages that require lexical pronouns in main and subordinate clauses.

A useful and lasting contribution made by Gilligan’s language classification is his recognition that within one language, variation does exist regarding when expletive subjects can be null and when they must be phonologically realised. This is reflected in his guideline definitions for classifying languages as restricted NSL or EXE-NSL, the observation that non-thematic subjects must be lexical under specific circumstances, and that they must be void of phonetic content under certain other conditions. This approach had not previously been given thorough and serious consideration; languages had normally been treated as either NSL or non-NSL. Gilligan was thus one of the earliest acknowledgers of language-internal null subject variation, which he represented in this work as an expletive pro-drop parameter.

The logical extension of Gilligan’s approach, then, is that we must now investigate under which linguistic settings thematic subjects are necessarily overt, and when they may be rendered null. We do not need at present to comment in depth on his description of core NSLs; Chapter 1 has already established clearly the common, erroneous assumption that subject pronouns in NSLs are never compulsory. A primary aim of the present research, then, is two-fold: first, to identify when subject pronouns are obligatorily present, and second, to fit these observations into a consistent framework, both language-internally and also cross-linguistically.

4 There appears to be some inconsistency in Gilligan’s classification: in his second survey, he identifies French as an NSL, based primarily on data relating to its subject clitics (1987: 213). We are not certain why Gilligan now designates French as a non-NSL in the current typology; he provides no defence or acknowledgement of this switch.
3.2.4 The role of parameters in describing the null subject

Gilligan may be one of the earliest scholars to question whether the cross-linguistic variation in null subject usage can be accounted for by parameters. Indeed, the linguistic community may have accepted by default that these differences could only be explained by parameterised settings. Gilligan introduces this as an assumption that should be examined thoroughly (1987: 29). While not committing himself to declaring that parameters are unable to explain the diverse treatments of null subjects in the world’s languages, he does suggest that this issue is at least worth investigating in depth before being accepted as an automatic axiom. This is one of the principal questions we attempt to address in our work. Do parameters provide an effective framework with which to explain the null subject phenomenon, either language-internally or cross-linguistically?

An additional claim that Gilligan puts forth, rather ambiguously, is that clitics and synthetic inflectional désinences, while certainly displaying differences with respect to their treatment of Agr (1987: 110), are similar enough to be considered as practically identical linguistic material for the purposes of feature marking and licensing. He writes that “the distinction between clitics and morphemes [should be] downplayed, since both forms contain the relevant features and arguably co-occur with null subjects” (Gilligan 1987: 185). This appears to be an idea developed quite haphazardly. Gilligan presents no systematic or formal argumentation in support of this statement. In fact, the data collected for this dissertation highlight that this is not the case. As we shall explore in further depth later on in this work, agreement by clitic and agreement by synthetic inflectional désinence (whether by verbal or adjectival agreement) produce different results. As a preview of what is to come, it will suffice to note here that European Portuguese will license a null subject in situations where gender is indicated by a clitic, whereas this cannot be achieved in Italian, although the latter is able to do so when gender is indicated by synthetic inflectional désinence. While there is some overlap, their effects are clearly not identical, and thus there is a strong case to be made that clitics and inflectional affixes are not equal agents in null subject licensing.
3.2.5 The influence of features

Another pertinent question that Gilligan raises, and one that does not appear to have been given full consideration, is the varying effects of specific features in allowing null subjects to be present. He notes that “it is widely recognized that not all agreement features interact equally with null pronouns. [...] There are three pronominal features which are encoded by agreement: person, number, and gender. The three are not all equal for purposes of identification, however” (Gilligan 1987: 228-229). Unfortunately, there is no systematic study of this in Gilligan’s research, nor does he allude to any scholar who has inspected the validity of this claim by a thorough and formalised examination.

He does present an observation that serves as the basis for a potential hypothesis, although there is little empirical follow-up on his statement and no theoretical argumentation to support such a claim. Gilligan, after observing that Romance verbs do just fine licensing null subjects without being able to mark gender in the standard verbal paradigm, concludes by deduction that the answer must lie in the other two features that are indicated, person and number. He contemplates that “given the fact that number agreement only appears to identify null pronouns in languages which also allow null pronouns without any agreement, I tentatively conclude that only the feature of person is necessary for the identification of null thematic pronouns” (Gilligan 1987: 234). Influenced in part by Gilligan’s claim, we follow up on his statement by investigating the effects of person-indicating morphology to see how valid this declaration is. The feature hierarchy that will be revealed at the end of Chapter 4 will show that, based on our data, person does appear to be the one feature of the three that would be most likely to act as the critical licensor of the null subject. That Brazilian Portuguese will tolerate null subjects when the person feature is used to disambiguate between competing antecedents, whereas the same cannot be said about number or gender, is evidence in support of Gilligan’s proposal. However, our data from Puerto Rican Spanish show that even person marking is not always enough to license the null subject. A further exploration of this topic follows in subsequent chapters.
This summary of Gilligan’s work should convince the reader that, while much has been added to our understanding of null subjects, such as the non-correlation of rich verbal morphology and null subject tolerance; the mutual independence of that-trace violations, subject inversion, and null subjecthood; and the existence of language-internal null subject variation, there is room for further expansion and amendment. Some of his points are addressed, with various levels of success, by scholars whose works we shall review throughout this chapter. Other hypotheses that he offers do not appear to have already been dealt with, and thus they serve as useful starting points from which to conduct further research. We have indeed chosen to pursue some of these leads in greater depth.

3.3 Sigurðsson (1993)

Sigurðsson confirms Gilligan’s assertion that there is no correlation between verbal agreement and the licensing of null subjects. He refers to work done by Hjartardóttir (1987) that argues that the ability to license pro was lost in the Icelandic language in the 18th and 19th centuries (1993: 248). The crucial offering is his observation that “this development did not relate to any weakening of verb inflection” (Sigurðsson 1993: 248-249), suggesting that verbal morphology had little to do with the licensing of pro in Old Icelandic. This, as a result, suggests that there must be some alternate theory to account for the inability in modern Icelandic to tolerate null thematic subjects.

The path pursued by Sigurðsson is not to discredit completely the role of identification by Agr in the licensing of null subjects, at least universally. Rather, Sigurðsson contends that in Icelandic, pro was never identified by rich Agr in the first place, a consequence of such an explanation being that this is consistent with the established information that null thematic subjects were lost in the language without any related alterations in the verbal paradigm (1993: 250). What is notable is that this avoids the need to disregard completely the influence of Agr on permitting null subjects. Instead, Sigurðsson has introduced a different parameter into the situation, one in which the
relevant settings are that rich Agr either has the ability to identify (a positive setting) or
does not have the ability to identify (a negative setting) referential pro. Modern Icelandic
carries the negative setting, Sigurðsson asserts, whereas positive settings are present in,
among others, Italian and Spanish (1993: 276).

If rich Agr is unable to license pro in Icelandic, then what does? Sigurðsson
claims that pro was identified under “free indexing with an NP in preceding discourse”
(1993: 250). He presents two arguments in support this. The first is that Old Icelandic
pro could not initiate discourse. Whereas in Italian, it would be possible to begin
discourse with a verb lacking a lexical subject, such as “(pro) Parlo” (‘I speak’), this was
not permitted in Old Icelandic (Sigurðsson 1993: 253). As pro was at the very beginning
of the discourse, there was nothing preceding that could act as a co-referring antecedent
to license it. The second piece of evidence is the extreme rarity of null subjects with non-
third person verbs, even though there are abundant and lengthy instances of direct speech
involving first and second person subjects in, for instance, Old Icelandic sagas
(Sigurðsson 1993: 253-254). This would be left unexplained by any theory assigning Agr
the ability to license pro. However, it would be consistent with a framework
incorporating narrative discourse topicality (Sigurðsson 1993: 254).

The significance of Sigurðsson’s claim for our purposes is that it is an example of
discourse factors having a systematic effect on the licensing of pro. Moreover, this
occurs in a language with rich verbal agreement, a noteworthy observation since
discourse considerations with regard to null subjects were normally applied principally to
those languages having little or no verbal agreement, such as in East Asian languages
(Huang 1984). However, Sigurðsson’s discovery also indicates that there are problems
with Gilligan’s purely syntactic approach to a cross-linguistic understanding of null
subject activity. What the Icelandic data and argumentation show is that the null subject
cannot be explained purely by morphosyntactic properties. Rather, we must also take
into account, within whichever framework we are applying to this problem, the relevant
non-syntactic considerations.

Another conclusion that Sigurðsson offers, that is germane to the investigations
attempted in the present work, is that the licensing of pro-drop, with respect to
identification in particular, can be explained by parametric variation (1993: 278). As
noted briefly above in the discussion on Gilligan’s work, it may not be the case that parameters serve as the optimal model by which to predict and explain null subject variation; we shall explore this in greater detail later on. Huang presents a similar conclusion for Chinese, arguing not only that pragmatics play a more dominant role in null subject license than syntax does, but also that these two linguistic areas are interconnected and need not be considered in mutually exclusive terms (1994: 259).

These works all encourage, or suggest the necessity of, a multi-disciplinary approach to the investigation of null subject variation across languages. They demonstrate that it is restricting and without tangible benefit to assume that the null subject can only be explained through one subfield of linguistics, whether this be syntax, pragmatics, discourse analysis, or even phonology. Rather, what is ultimately required is a solution that incorporates all of these fields into an accommodating framework. It is with this aim as our motivation that we now turn to a review of more recent works on null subject variation, which guide the approaches taken in the present research.

3.4 Samek-Lodovici (1996)

3.4.1 Using Optimality Theory to understand syntax

Up until the mid-1990s, the vast majority of attempts to explain cross-linguistic variation in null subject usage adopted a largely syntactic framework. This began in the early 1980s with the Principles and Parameters Theory (Chomsky 1981; 1986), where inviolable principles were established that reflected similarities across all languages, while the parametric component took into account inter-linguistic differences. This approach was refined through the Minimalist Program (Chomsky 1991; 1992; 1995; Chomsky and Lasnik 1993). In Minimalism, the mechanism that obligatorily checks features is the universal component, and the dissimilarities between languages emanate from whether selected features are designated as strong or weak in each specific language. These, in turn, are contained in the lexicon, so cross-linguistic variation is a result of different lexica.
Starting from the mid-1990s, we have increasingly witnessed syntactic explanations within a new framework. This model, Optimality Theory (Prince and Smolensky 1993), consists primarily of three components: GEN (for “generator”), which creates a potentially infinite number of inputs to be evaluated; EVAL (“evaluator”), which analyses each input item; and CON (“constraints”), the universal set of constraints which EVAL utilises when it examines the inputs. Once it has accomplished this task, EVAL then outputs the candidate deemed to be optimal, and by definition, grammatical. This output is decided upon in each language by its unique ranking of the constraints. So linguistic universality is modelled as the set of constraints common to all languages, while variation among languages is determined by their singular ordering of these shared constraints.

Samek-Lodovici, following other like-minded linguists such as Grimshaw (1993; 1995) and Legendre (1996), extended Optimality Theory’s applications to syntax, where previously it had been used primarily to investigate phonological questions. In comparing the feasibility of using Optimality Theory to study null subject usage, Samek-Lodovici asserts that it compares rather favourably with the two traditional syntactic frameworks. He argues for the advantage of Optimality Theory over Principles and Parameters first by addressing the latter’s inherent nature. The outlook offered by Principles and Parameters is a necessarily exclusive one, while that provided by Optimality Theory is all-encompassing, and it is this universality that makes the second model more appealing. Samek-Lodovici notes that:

“[…] once the value of an hypothetical parameter specifying the direction of focus-alignment as either leftward or rightward is set, the opposite value becomes inaccessible. In the Optimality Theoretic perspective, on the other hand, all constraints are universal, and therefore they are present in the grammar of every language. Thus, if there exist a leftward and a rightward version of the abstract ALIGNFOCUS constraint, they should both be part of each language’s grammar.”

(Samek-Lodovici 1996: 155)
Optimality Theory suggests that parameters are an expendable component to a model of grammar, since parameter variation can now be accounted for through constraint rankings (Samek-Lodovici 1996: 165).

A further improvement that Optimality Theory provides over a Principles and Parameters framework is that the former can successfully account for intra-linguistic mixed patterns, such as overt subject pronoun availability in Hebrew (cf. the approaches in Borer 1983; 1989) and focusing in the Chadic language Kanakuru, which Tuller (1992) was unable to explain using Principles and Parameters, as Samek-Lodovici (1997: 167) points out. The flexibility offered by Optimality Theory in this respect is one of our strongest motivations for considering such an approach to Romance subject pronoun usage.

In relation to the Minimalist Program in particular, Samek-Lodovici asserts that Optimality Theory is no worse a framework. In fact, the two models are more similar than their surface appearances would initially indicate. Their core resemblance is that both are based upon the concepts of selectively permissible violation and relative ranking. Samek-Lodovici, alluding to Minimalism’s “Last Resort” principle, which prevents the expensive operation “Movement” to apply unless the derivation would otherwise crash (Chomsky 1993: 32), remarks that this is equivalent to a situation in Optimality Theory whereby we disobey a constraint against movement in order to appease a higher-ranked constraint requirement on feature checking (1996: 282). Both Optimality Theory and Minimalism allow for violations to occur so that more urgent operations and higher-priority constraints take precedence over those that are not as crucial. In this respect, these two frameworks actually do complement each other quite well.

Although Samek-Lodovici makes a convincing argument supporting Optimality Theoretic applications to syntax, some will question why we should disregard Minimalism in favour of Optimality Theory, if the two are of largely equal competence. We explore one possible advantage of Optimality Theory over the Minimalist Program when we discuss Margaret Speas’ (1997) contribution below.
3.4.2 Questioning pro

Samek-Lodovici proposes another stance that conservative linguists may find shocking. He denies the existence of pro, proexp, and PRO, as outlined by Chomsky (1981; 1982), maintaining that structures incorporating these grammatical units “are actually structures lacking a structurally realized subject” (1996: 7). His assertion is based on the result that both lexical subject pronouns and pro equally satisfy his proposed constraints SUBJECT and PARSE, because pro, although it is silent, still exists as a syntactic, structural entity (Samek-Lodovici 1996: 75). This situation would thus not lead to a language-internal or cross-linguistic explanation of null subject pronoun variation. The question that arises, naturally, is whether we need to re-evaluate the existence of these long-accepted phonologically null constructions, or whether Samek-Lodovici’s two constraints themselves require reworking to be compatible with all versions of pro and PRO. We expand further on this discussion after we have identified and described the relevant constraints in greater detail below.

Samek-Lodovici (1996: 81) offers on behalf of his stance an appeal to the concept of structural deficiency, as proposed by Cardinaletti and Starke (1994), who divide all pronominals into three categories: strong, weak, and clitic, in increasing order of structural deficiency. This is accomplished on the basis of, among other criteria, word stress; the stronger a pronominal, the more amenable it is to being vocally emphasised.

Cardinaletti and Starke suggest that there is a correlation between structural deficiency and referential dependence. Strong pronominals may refer to entities that have yet to appear in the discourse, whereas the weak and clitic groups must refer to discourse prominent entities, in other words, those that have already been mentioned previously in the discourse. So a direct relationship has been established between the level of structural deficiency and the type of antecedent available for co-reference.

Samek-Lodovici disagrees with the proponents’ assessment of the null subject, which they represent as pro, as a weak pronominal, arguing on solid grounds that this would imply that null subjects are structurally stronger than clitics. If word stress is one of the criteria judging structural deficiency, then it is indeed puzzling how null subjects could be considered stronger than clitics. Apart from that piece of contention, Samek-
Lodovici finds Cardinaletti and Starke’s framework otherwise suitable. Null subjects, having no phonetic substance, are the weakest pronominals, and so their high structural deficiency implies a comparable level of referential dependence. This is consistent with Samek-Lodovici’s findings that null subjects must have topic antecedents, which are the most discourse prominent elements available for referral. The assumption that pro is structurally unrealised, then, is consistent with at least one independent framework, a beneficial factor to Samek-Lodovici’s position on pro. Other authors have taken a similar stance on the non-existence of pro, such as Alexiadou and Anagnostopoulou (1998) and Manzini and Savoia (1997; 2002), inter alia.

Samek-Lodovici also disputes the existence of pro$_{expl}$ by attempting to show that it is subject to requirements that cannot all be satisfied simultaneously. In (7) below, we see that pro$_{expl}$ heads a pro$_{expl}$-chain and expect it to be able to bind an anaphor in its scope, but in fact it is unable to bind anaphoric se stesso, resulting in the ungrammatical structure:

(7) *Sembrava a se stesso non guadagnare abbastanza nessuno.
   [IP pro$_{expl}$ sembrava [a se-stesso] [IP ti non guadagnare abbastanza nessuno]]

   seemed [to himself] [ not to-earn enough anybody]]

   ‘Nobody seemed to himself to earn enough.’

   (Samek-Lodovici 1996: 264)

If we acknowledge that an acceptable sentence gets produced if we either occupy [Spec, IP] with a lexical subject that can bind se stesso, as in (8), or delete the anaphor, as in (9), then the only possible rescue from this paradox is to assert that pro$_{expl}$ does not have the ability to bind, cannot carry phi-features, and is non-referential:
(8) Nessuno sembrava a se stesso guadagnare abbastanza.
[IP Nessuno; sembrava [a se-stesso] [IP tì guadagnare abbastanza.]]

[ Nobody seemed [to himself] [ to-earn enough ]]  

‘Nobody seemed to himself to earn enough.’

(9) *Sembrava non guadagnare abbastanza nessuno.
[IP pro expl sembrava [IP tì non guadagnare abbastanza nessuno]]

[ seemed [ not to-earn enough anybody]]

‘Nobody seemed to earn enough.’

(Samek-Lodovici 1996: 265)

However, if we are to make this presumption about pro expl, then a sentence such as (10), which clearly shows that pro expl does bear phi-features to trigger agreement with the matrix verb sembrano, should not be grammatical. (10) is, in fact, legitimate:

(10) Questa volta, sembrano non aver votato molti elettori.
[Questa volta] [pro expl sembrano [tì non aver votato [molti elettori]]]

[This time, seem-pl [ not to-have voted [many voters]]]

‘This time, few voters seem to have voted.’

(Samek-Lodovici 1996: 266)

Samek-Lodovici concludes then, in view of the above data as well as other arguments, that pro expl is not structurally realised. As for PRO, he does not present any direct justification against its existence, so it must be assumed that the arguments he proposes to support his denial of pro should be appropriately applied in a comparable analysis of PRO.
In light of our observations in Chapter 1 on the proposed lexical subjects of infinitives in Caribbean Spanish, and the resulting ambiguity as to whether these units are pro or PRO, the tenability of Samek-Lodovici’s strongly argued position must be given thorough investigation. We shall briefly return to this issue later on, although without attempting a complete resolution of it, merely indicating points relevant to our research and leaving the finer details of the debate to future research. In the meantime, the terminology we adopt in the present work follows the traditional assumption of the null subject being represented by pro, consistent with those whose works we review below.

### 3.4.3 Constraints proposed

Samek-Lodovici formulates four constraints to account for null subject employment. They are:

i) SUBJECT
Fulfilled when the highest A-specifier of a clause is structurally realised

(Samek-Lodovici 1996: 25)

This merely specifies that the [Spec, IP] position must be materially occupied. SUBJECT is evocative of Chomsky’s Extended Projection Principle (1982: 10); this is the Optimality Theoretic equivalent of the requirement that all clauses must have subjects.

ii) PARSE
Fulfilled when all input items appear in the output

(Samek-Lodovici 1996: 29)

PARSE appears to be a partial expression of the standard faithfulness constraint. It is a unidirectional version of the FAITH constraint offered by most Optimality Theoreticians,
specifying only that all input items must appear in the output. Note that this constraint is not directly violated in the other direction; when the output contains extraneous material not present in the input, there is no infringement on PARSE.

iii) FULL-INT
Fulfilled when lexical material is interpretable

(Samek-Lodovici 1996: 30)

This constraint oversees the interpretation of all projections, and observes a violation whenever overt projections have not been theta-assigned. It may be disobeyed more than once in a sentence, for example, if it contains both an instance of do-support (as explored by Grimshaw (1993; 1995)) and expletive elements.

iv) DROPTOPIC
Fulfilled when arguments with a topic antecedent are left unrealised

(Samek-Lodovici 1996: 31)

DROPTOPIC is Samek-Lodovici’s innovative contribution to the exploration of null subject usage with an Optimality Theoretic framework, and is the constraint most applicable to our study. We anticipate that DROPTOPIC will be the principal component of any thorough analysis of null subjecthood, and hence this constraint, as well as revised versions thereof proposed in subsequent works, is the one that we shall examine with the deepest scrutiny.

One claim that Samek-Lodovici repeatedly declares, and which serves as a consistent theme throughout his study, is the direct correlation between null subjects and topic antecedents. He maintains that null subjects must be licensed by a topic subject, and that a non-topic antecedent requires a co-referential overt pronominal subject, as summarised by the following data:
(11a) *Questa mattina [il fratello di Gianni] ha visitato la mostra. Nel pomeriggio *e, /
??egli / lui ha visitato l’università.*

(11b) *Questa mattina [il fratello di Gianni] ha visitato la mostra. Nel pomeriggio e, /
??egli / *lui ha visitato l’università.*

(Samek-Lodovici 1996: 44-45).

While the sentences above are consistent with the grammatical judgements offered by the native speakers of Italian consulted in collecting the data for the present work, and thus support the link between null subjects and topic subject antecedents, we have reason to believe that this is not a linguistic universal. In particular, we shall see in the next chapter that a parallel sequence in Catalan will permit the null subject to be coreferential with either the topic or non-topic antecedents; we leave this issue here for the moment.

Another concern arising from Samek-Lodovici’s assertion that is not explicitly addressed at first is how a discourse-initiating utterance may begin with a simple first or second person verb without a subject pronoun, such as parlo (‘I speak’) and parli (‘you speak’), and to a more questionable extent, a third person parla (‘he speaks’), when there is no preceding topic to serve as an antecedent. Samek-Lodovici elaborates on the case involving first and second person by proposing a deictic component to his constraint. He proposes that the speaker and hearer always act as deictic topics, and therefore a situation where a pronominal subject representing a discourse participant lacks a non-topic antecedent never occurs. This can be codified formally by splitting DROPTOPIC into two independent constraints: DROPTOPIC_deic, which accounts for differences between third and non-third grammatical persons, and DROPTOPIC_disc, which we assume to be identical to the initial DROPTOPIC constraint, dealing with discourse topics only (Samek-Lodovici 1996: 102-103). As for the situation with a subject pronoun-lacking third person verb, Samek-Lodovici offers a parallel explanation. Discourse-initial parla is much more acceptable when accompanied by a deictic extralinguistic sign, such as finger-pointing (Samek-Lodovici 1996: 195). We expand on Samek-Lodovici’s
examination of the relationship between third-person and non-third person null subject availability in the chapters following.

Returning briefly to DROPTOPIC, we claim that it will indeed require some fine-tuning. The incompatibility of this constraint has already been pointed with the situation in Catalan, and as we shall see in the next chapter, some Romance varieties will tolerate a null subject with a topic antecedent when there is an intervening clause. It is not certain at this point whether resolving this inconsistency involves a reformulation of the DROPTOPIC constraint itself, or whether a better proposal would be to propose an additional constraint to interact independently with DROPTOPIC. Either prospect would have to take into account, and be harmonious with, both the results offered by Samek-Lodovici and those introduced in the present work.

We now proceed to an evaluation of how these constraints interact, the outcomes that result from their ordering, and how the relative rankings of the constraints explain cross-linguistic variation.

3.4.4 Constraint interactions, their rankings, and cross-linguistic variation

We have previously mentioned in passing how the different ordering of constraints reflects the various methods by which languages treat the same linguistic entity. A more comprehensive review is pursued here.

It would be beneficial to investigate first the ramifications of constraint rankings. Consider the possibilities offered by alternately assigning priority to Samek-Lodovici’s DROPTOPIC and PARSE. If we rank the former above the latter, then an input containing the null subject will be selected as optimal, whereas if PARSE takes precedence over DROPTOPIC, then the sentence with the overt subject will emerge in the output. This is encapsulated in the two charts below:\footnote{We do not go into the details of the candidates involving the postverbal subject and the null structure, as they are beyond the immediate scope of this exposition and the focal areas of this dissertation.}

5
In (12a), the null subject candidate in row (c) contravenes PARSE once by virtue of not overtly representing the input Gianni. However, the preverbal subject candidate in row (a) commits a more serious violation of the higher-ranked DROPTOPIC by allowing lui, co-referential with topic antecedent Gianni, to appear lexically. The null subject candidate, although it violates PARSE, does duly observe the higher priority DROPTOPIC constraint and emerges as the optimal result of the two.

Similar argumentation accounts for the opposite effect in English. The fundamental difference between English and Italian is their ordering of the two relevant constraints. In English, it is PARSE that is the higher-ranked constraint. Therefore, in (12b), the overt subject candidate (a), while violating DROPTOPIC for the same reason that its Italian correspondent did above, does obey the higher-ranked constraint PARSE, whereas its null subject counterpart in (c) violates it. In English, the null subject is not permitted in this instance as a result of PARSE ranking above DROPTOPIC.
reverse is true, as in Italian, the null subject is licensed. Notice that applying the Italian constraint ranking to English would produce the null subject candidate (c) as the optimal candidate, an undesirable result. This shows compellingly that it is the language-specific ordering that accounts for the inter-linguistic differences.

We do not need to comment here extensively on FULL-INT, which is satisfied equally by all four candidates in the chart and therefore makes no contribution to our attempt to discriminate between the options. As for SUBJECT, we see that it is rendered ineffective in the analysis above because all candidates that violate it also violate a higher ranked constraint, which is a more severe action.

We hope to have convinced the reader of the tenet that varying constraint rankings account for cross-linguistic differences. We summarise below the observations presented above, and extrapolate on how an Optimality Theoretic framework could complement typological analyses as well:

**Italian:**

DROPTOPIC >> PARSE >> SUBJECT (because null thematic subjects are permitted)
FULL-INT >> SUBJECT (because overt expletive subjects are prohibited)

Constraints could be ranked in Italian as follows:

DROPTOPIC >> PARSE >> FULL-INT >> SUBJECT
DROPTOPIC >> FULL-INT >> PARSE >> SUBJECT
FULL-INT >> DROPTOPIC >> PARSE >> SUBJECT

**English:**

PARSE >> DROPTOPIC
SUBJECT >> FULL-INT
PARSE >> FULL-INT (because null thematic subjects are prohibited)
SUBJECT >> FULL-INT (because overt expletive subjects are required)

---

6 At least, under the specific syntactic conditions examined in the review presented here. The statement that "null subjects are permitted in Italian" is, of course, a broad generalisation, and is anathematic to our underlying ethos of language-internal variation in null subjecthood. The simplification here is intentional, however, in order to present without disturbance the basic arguments supporting language-specific constraint rankings and their typological applications.
So possible orderings of these constraints for English would include:

PARSE >> DROPTOPIC >> SUBJECT >> FULL-INT  
SUBJECT >> PARSE >> DROPTOPIC >> FULL-INT  
PARSE >> SUBJECT >> DROPTOPIC >> FULL-INT

The actual ranking among all of these potential candidates is eventually determined, of course, after further investigation of other English and Italian data involving these constraints.

By extension, we could also imagine that constraint rankings exist for the other two combinations of thematic and expletive subject pronoun availability. It would be a reasonable hypothesis to state that a language with overt thematic subjects but null expletive subjects could be characterised by rankings observing the patterns below:

PARSE >> DROPTOPIC (overt thematic subjects)  
FULL-INT >> SUBJECT (null expletive subjects)

Languages with null thematic subjects but permitting overt expletive subjects (e.g., some southern Italian dialects, Galician) would obey the following restrictions:

DROPTOPIC >> PARSE >> SUBJECT (null thematic subjects)  
SUBJECT >> FULL-INT (overt expletive subjects)

Thus: DROPTOPIC >> PARSE >> SUBJECT >> FULL-INT

One potential application of constraints and their rankings to any work involving typologies is that, if we notice an unattested language type, it may be possible to examine the constraint rankings predicted for that group to determine whether such the category under question is impossible, or whether it is plausible but so far undocumented. If an analysis can show that a specific ordering of constraints is impossible or inherently contradictory, this may well be conclusive evidence against the existence of such a language type.
3.4.5 Points of further research

Samek-Lodovici’s work raises numerous opportunities for additional study, either by intentionally setting an issue aside for future scholars to develop, or by proposing arguments that have not been explored in satisfactory depth and rigour, thereby leaving them still open to additional scrutiny. We have already indicated a couple of them above, such as the uncertain existence of pro and its phonologically null cousins in Section 3.4.2, and the questionable absoluteness of null subjects requiring topic antecedents as exposed in Section 3.4.3. A few more issues are identified below.

It is worth pondering the exact relationship between discourse participants and grammatical third persons. This was briefly alluded to in Section 3.4.3, when Samek-Lodovici asserted that the differences in topic antecedence between first and second person subjects, and that of third person subjects, could be explained by deictic considerations. Further, he mentions that a fundamental contrast between the two groups is that third person pronominals are accompanied by “obligatory strong usage” (1996: 102). This characterisation is tinged with the claim that such pronouns are only used for emphasis or contrast, and it has already been well established in the present work that this is indeed not the case. While he presents limited observations in support of this claim, Samek-Lodovici does not pursue a systematic investigation of this with the thorough treatment he affords on the other syntactic properties he examines in his study, such as subject inversion and case assignment. Aside from proposing a split between DROPTOPIC_{deic} and DROPTOPIC_{disc}, there is no standardised attempt to incorporate this split within an Optimality Theoretic framework. This is one area of research that the present work will pursue through formal procedures.

Although he does not look rigorously at differences in null subject licensing between first and second persons, and the third person, Samek-Lodovici does analyse the person feature in context with the other two features normally associated with verbal agreement morphology, number and gender. He presents data in support of the hypothesis that when feature agreement present in Spec-head agreement gets lost in agreement under c-command, the feature least likely to disappear is person (1996: 172).
He cites Saccon (1993) and Brandi and Cordin’s (1989) examples in Fiorentino and Trentino as an indication:

(13) \textit{La Maria la parla.}\hfill (Spec-head agreement)
\begin{itemize}
\item The.3Fs Mary.3Fs cl.3Fs speaks.
\item ‘Mary speaks.’
\end{itemize}

(14a) \textit{Gl’è venuto la Maria.}\hfill (c-commanding agreement, with loss of gender agreement)
\begin{itemize}
\item There is.3s come.3Ms the.3Fs Mary.3Fs.
\end{itemize}

(14b) \textit{*L’è venuta la Maria.}\hfill (Samek-Lodovici 1996: 289)
\begin{itemize}
\item cl.3Fs is.3s come.3Fs the.3Fs Mary.3Fs.
\end{itemize}

‘Mary arrived.’

(15a) \textit{Gl’è venuto delle ragazze.}\hfill (c-commanding agreement, with loss of number and gender agreement)
\begin{itemize}
\item There is.3s come.3Ms some.3Fp girls 3Fp
\end{itemize}

(15b) \textit{*Le sono venute delle ragazze.}\hfill (Samek-Lodovici 1996: 289)
\begin{itemize}
\item There is.3s come.3Ms some.3Fp girls 3Fp
\end{itemize}

‘There arrived some girls.’

Where indication of number and gender may be eliminated or neutralised, person is the most resilient feature, generally preserved regardless of whether the agreement configuration is Spec-head or c-command\textsuperscript{7}. This is at least consistent with Gilligan’s scantily-supported observation in Section 3.2.5 above that the person feature was the most critical of the three. We do note, however, that all the data that Samek-Lodovici

\textsuperscript{7} But see Ledgeway (2004) for counterexamples in the dialect of Ripatransone in central Italy.
provides to assert this claim involve third-person subjects. The argumentation would have been more convincing if examples involving the same phenomena with first and second plural subjects had been included as well. The present research will review the relevant strengths of the three agreement features by systematic methods.

Samek-Lodovici also makes a quick prediction that “null subjects should [...] always show the fullest agreement paradigm available in that language, even when the language allows for agreement loss” (1996: 193). While offering that, as far as he is aware, this statement is not contradicted by any language with null subjects displaying weaker agreement than what is available to lexical subjects occupying the [Spec, IP] position, he does not seek a proper formulation of this generalisation. Melvyn Cole (2000) proposes a hypothesis that seeks to address certain aspects of Samek-Lodovici’s assertion within a formal framework; we survey the success of this attempt later in this chapter.

Finally, Samek-Lodovici demonstrates convincingly in his work that the option of null subjects always being optional is a misguided generality. However, he makes a claim that may be too strong on the other extreme. Samek-Lodovici contemplates the “complementary distribution of null and overt subjects in pro-drop languages” (1996: 96). It is one thing to say that null subjects are not always optional; it is quite another to assert, on the other hand, that they are never optional, as his statement implies. We must allow for the possibility that in certain syntactic circumstances, the use of a lexical (or null) subject pronoun is available but not strictly required. It may be rash to assume that only a null or only a lexical subject pronoun is able to appear in any particular syntactic environment. Our review of Baković (1997) below discusses this in greater detail.

3.5 Speas (1997)

Additional evidence to support Samek-Lodovici’s revelation of the compatibility of syntax with an Optimality Theoretic framework is presented by Speas (1997), who further explores the potential success of accounting for cross-linguistic null argument
variation outside of the traditional treatments normally offered by the Principles and Parameters frameworks.

Before introducing her innovative approach to this issue, Speas first addresses a potentially problematic dilemma that Samek-Lodovici does not touch upon. It seems radical to offer as an alternative to Principles and Parameters, a theory predicated on an initial assumption that syntax is composed of inviolable principles, a framework such as Optimality Theory, which contends axiomatically that no constraint is immune from contravention. Such a drastic and opposite approach to the traditional standard is quite unexpected, and initially comes across almost as counterproductive and paradoxical to any attempt at furthering our understanding of how language functions.

Speas suggests that all syntactic constraints are indeed violable (1997: 171), and points out that the generally-held belief of syntactic principles operating without exception is not sustainable. She mentions that “every principle of PPT contains some hedge, some special clause to cover cases which do not obey a simple version of the principle” (Speas 1997: 180). In other words, hedges are merely ad hoc patches applied, as appropriate and advantageous, to syntactic rules in order to turn exceptions discrediting the principle into “non-exceptions” that fall under a wider, revised principle. It is a substantial drawback of any theory to have a weakening of its components in an inconsistent fashion. Speas contends that this watering-down of otherwise solid principles would be unnecessary if we adopted a framework allowing for violations in constraints, and that such a move would render superfluous the need for all these exception-granting provisos (1997: 171-172), thus producing a more watertight and defensible theory.

Speas initiates her treatment of syntax via an Optimality Theoretic approach by first considering head-complement order. She proposes a constraint called SATISFY, fulfilled if morphosyntactic features are checked in a specifier position, in addition to Grimshaw’s (1997) already-established STAY, and maintains that it is the relative ranking of these two constraints that account for the differences between English and Japanese head-complement order (Speas 1997: 176-7). In English, STAY outranks SATISFY, whereas the opposite is true in Japanese. The benefit of this approach to head-complement order is that “ranking the constraints does not simply replace GB
parameters or the MP notion of feature strength; it also allows the constraints to be stated in [the] simplest way possible. If constraints were not violable, they would contain extra stipulations analogous to parameters or feature strength specifications” (Speas 1997: 177). That is to say, instead of having a dual component theory, as in universal principles and language-varying parameters in a Government and Binding approach, or language-determining strong and weak features in addition to functional operations as in Minimalism, we have one set of language-universal constraints in Optimality Theory, and it is just the manner in which those constraints are ranked relative to each other that account for cross-linguistic variation. Speas argues therefore that the benefit of an Optimality Theoretic framework over Principles and Parameters is its efficiency and economy in representation. A complete investigation of this question will not be attempted in this work. If insights directly pertaining to the present research and the data presented therein are relevant to this debate, they will be mentioned as appropriate. We set a comprehensive review of this issue aside for further research elsewhere.

The compatibility of Optimality Theory to head-complement order having been established, Speas then proceeds to account for null arguments in finite clauses, using the same framework. She identifies three constraints pertinent to subject and object pro:

i) CONTROL
   Fulfilled if a null pronoun is controlled in its control domain

ii) FREE PRONOUN
    Fulfilled if it is free in its governing category

iii) MAX(PRO)\(^8\)
    Fulfilled if Pro\(^9\) occurs in the input as well as the output

(Speas 1997: 189)

---

\(^8\) This constraint seems identical to what Archangeli, among many others, would call faithfulness, normally represented as FAITH (1997: 11). Speas does not explain why she chooses to name her constraint MAX(PRO), and there do not appear to be any discernible differences between FAITH and MAX(PRO).

\(^9\) Speas, in accordance with Huang (1984) and Borer (1989), chooses not to differentiate between PRO and pro, treating the two as one and the same, calling this entity “Pro” (1997: 182, footnote 8). We do not consider the merits of their approach here, leaving it for future research.
Whether a language allows subject and/or object pro in finite clauses depends on how it orders these three constraints. A summary table is offered below:

(16)

<table>
<thead>
<tr>
<th>Language</th>
<th>Subject pro</th>
<th>Object pro</th>
<th>Account</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thai, Korean</td>
<td>yes</td>
<td>yes</td>
<td>FREEPRN, MAX(PRO)&gt;&gt; CONTROL</td>
</tr>
<tr>
<td>English</td>
<td>no</td>
<td>no</td>
<td>CONTROL, FREEPRN &gt;&gt; MAX(PRO)</td>
</tr>
<tr>
<td>Spanish</td>
<td>yes</td>
<td>no</td>
<td>CONTROL, FREEPRN &gt;&gt; MAX(PRO); subject pro is controlled by Agreement or by higher subject</td>
</tr>
</tbody>
</table>
| Mandarin | yes | no | ????

(Speas 1997: 197)

Speas maintains that this typology results from the interaction of the three constraints in question. If both FREE PRONOUN and MAX(PRO) outrank CONTROL, then both argumental pros are permitted; otherwise, subject pro and/or object pro is not tolerated. She is unable to cite any attested languages that license object pro but not subject pro.

It should be rather clear that the details of her analysis on this particular topic are a bit problematic. For instance, the ordering of constraints for English, and for Spanish and Mandarin, is identical, even though their tolerances for subject pro are clearly different. Speas discriminates between the two by inserting an amendment for Spanish and Mandarin next to their constraint ranking that “subject pro is controlled by Agreement or by higher subject”. This is surely an undesirable outcome of her analysis. This move appears to be merely an ad hoc explanation external to the Optimality Theoretic model being used, and this renders her earlier opposition to hedges in traditional syntactic frameworks as an inconsistent one.

---

10 In a subsequent review, Speas does assert that this is impossible, because object pro is unable to satisfy simultaneously both CONTROL and the constraint BTB, fulfilled if a pronoun is free in its Binding Domain (2001: 414).
It is granted that this shortcoming both undermines the strength of her argument that syntactic matters may always be accounted for through Optimality Theory, and also leaves many questions unanswered on how precisely to formulate an explanation of cross-linguistic subject *pro* variation using a system of violable constraints. However, Speas' work is indeed valuable to us because it does demonstrate the potential to account for null subject differences across languages using Optimality Theory. The major step that is lacking in her presentation is a considerable refinement and sharpening of her model, but the possibility of doing so successfully is foreseeable.

A second, more significant, drawback to Speas' model is that it is unable to account for variation within any one particular language. Ideally, we would like one unitary model that would be able to accommodate both intra- and inter-linguistic variation. While we concede that Speas' basic approach, once it is constructed accurately, should be able to provide a thorough treatment of null subject variation between languages, it still appears unable to explain differences in null subject usage within the various constructions of a single language. This is a crucial criticism because, as we shall see clearly in the following chapter, a certain language may tolerate a null subject in a particular syntactic environment but not in a different one. The implication in Speas' approach is that languages either do have, or do not have, subject (and object) *pro*. Indeed, this position was one of our primary motivations for arguing against a single, all-encompassing "null subject parameter"; it is simply not informative or accurate to say that a language is either [+pro-drop] or [-pro-drop]. The situation, as we realise by now, is not that clear-cut. She neglects Samek-Lodovici's contributions showing, for example, overt and null subject pronoun variation within Italian itself (1996: 44-45). Melvyn Cole alludes to this problem in Speas' framework, writing that it:

"[...] deals with null subjects as if they could occur right across a language in which they do occur, irrespective of circumstances. Such a view can no longer be upheld in the light of the data [...] which [indicate] that very individual circumstances involving morphology and, in particular, pragmatics are necessary for a null subject to occur in any language."

(Cole 2000: 209)
We shall examine a bit later how Cole goes about addressing the deficiencies in Speas’ model, and to what extent he is successful in doing so.

### 3.6 Baković (1997)

One potential drawback of an Optimality Theoretic approach is the incorporated assumption that the optimal candidate outputted by the framework mechanism is the only grammatical one (Prince and Smolensky 1993: 2). While this may work fairly well for phonology, the linguistic field to which Optimality Theory was originally and most widely applied, it does not appear to be as compatible with syntactic issues. Indeed, there are instances when more than one syntactic representation may be comparably acceptable, with no nuanced difference whatsoever separating them. The problem that thus arises is how to reconcile a system that, at least in its most basic form, prefers a solution where only one optimal candidate is selected, with the linguistic reality that there are occasions where true optionality between alternate representations needs to be accounted for. An ideal answer will refine the framework in such a way as to make Optimality Theory correctly predict, whenever the situation is fitting, all equally grammatical results produced in the output.

Baković attempts to resolve this discrepancy by investigating the case of optional English complementiser *that* in embedded clauses, and fitting an Optimality Theoretic account to explain it. He first demonstrates instances in relative and complement clauses where the overt expression of the complementiser is truly optional:

(17a) The coat (that) he always wears doesn’t fit him.

(17b) I think (that) the coat doesn’t fit him.  

(Baković 1997: 1)
Baković points out, however, that there are situations where the appearance of *that* is not subject to free variation. It must be overt where the subject has been extracted from a relative clause or when there is adjunction to a complement clause:

(18a) The coat *(that) doesn’t* fit him might fit me.

(18b) I think *(that) on him, no coat looks good.  
(Baković 1997: 2-3)

There are also circumstances dictating that the complementiser must not be expressed, such as when the subject has been extracted from a complement clause:

(19) Which coat do you know *(that) doesn’t* fit?

(Baković 1997: 3)

To account for those data, Baković proposes a feature for subordination, called [SUB], that marks CPs as [+SUB] and IPs as [-SUB]; whenever an embedded clause is inputted into the OT architecture, it carries the appropriate specification for [SUB]. Baković then introduces the constraint FAITH[SUB]:

(20) FAITH[SUB]
Fulfilled when the input and the output values of [SUB] are identical

(Baković 1997: 3)

This, coupled with various manifestations of the markedness constraint MARK, as proposed by Grimshaw (1997), is sufficient to explain the requirement, permissibility, or prohibition of an overt *that* complementiser\textsuperscript{11}. We saw in the previous subsections that

\textsuperscript{11} We choose not to recapitulate here, for reasons of space and scope, the fact that the constraint MARK is itself not irreducible, but instead is justified by and based on more primitive syntactic considerations, such
examples of syntactic variation treated by Optimality Theory result from different constraint rankings; it should not come as a surprise then that the range of different possibilities in English that expression is realised by the specific ordering of the pertinent constraints.

Baković's analysis is as follows. Complementisers are optional when FAITH[SUB] outranks MARK; it is fatal for the output not to match the input, even at the expense of producing a marked structure\(^\text{12}\). Therefore, if the linguistic string inputted contains a CP, then that will be the output as well. If it has an IP, then an IP will be present after being analysed by Optimality Theory. In brief:

(21)

Candidate (a): CP: The coat [CP that he always wears t] doesn't fit him.
Candidate (b): IP: The coat [IP he always wears] doesn't fit him.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a. #* CP</td>
<td></td>
<td>(*)</td>
<td></td>
</tr>
<tr>
<td>b. IP</td>
<td>*!</td>
<td>(*)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>osteness</th>
<th>Input: [-SUB]</th>
<th>FAITH[SUB]</th>
<th>MARK</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. CP</td>
<td>#!</td>
<td>(*)</td>
<td></td>
</tr>
<tr>
<td>b. #w IP</td>
<td></td>
<td>(*)</td>
<td></td>
</tr>
</tbody>
</table>

(Baković 1997: 3)

The left table above shows what happens when candidate (a) with the CP is inputted, while the right table has candidate (b) as its input, and demonstrates how it gets chosen.

The argumentation accounting for obligatorily present or absent that complementiser proceeds similarly, except with the constraints now in the opposite order. Complementiser that is required when the constraint MARK-IP (indicating that IP is the non-neutral constituent) outranks FAITH[SUB]; in this case, Optimality Theory will

\(^{12}\) For this part of the analysis, it is not relevant whether the marked structure is IP or CP, since it is rendered inconsequential by MARK being outranked by FAITH[SUB].
select the output candidate that lacks IP, even if it means being unfaithful to the input and violating the lower priority FAITH[SUB] constraint:

(22)

Candidate (a): CP: The coat [CP that *t doesn’t fit him] might fit me.
Candidate (b): IP: The coat [IP *t doesn’t fit him] might fit me.

| Complementizer obligatoriness: MARK-IP » FAITH[SUB] |
|---------------------------------|------------------|------------------|
| a. [CP] | | | a. [IP] |
| b. [IP] | *! | * | b. [IP] | *! |

(Baković 1997: 4)

The two tables above show that, regardless of the type of embedded clause entered as input, the Optimality Theoretic system will also indicate the structure with the CP as the preferred one, since it does not violate the fatal MARK-IP constraint.

Finally, the absence of the complementiser *that* is specified when the constraint MARK-CP outranks FAITH[SUB]. The optimal candidate chosen will be the one lacking the complementiser, the one containing IP, because it obeys the higher-ranked MARK-CP constraint, at the expense of a less severe violation, unfaithfulness to the input as represented by FAITH[SUB]. The table below summarises this result:

(23)

Candidate (a): CP: Which coat do you know [CP that *t doesn’t fit]? 
Candidate (b): IP: Which coat do you know [IP *t doesn’t fit]?

| Complementizer absence: MARK-CP » FAITH[SUB] |
|---------------------------------|------------------|------------------|
| **Input**: [,+SUB] | MARK-CP | FAITH[SUB] | **Input**: [−SUB] | MARK-CP | FAITH[SUB] |
| a. [CP] | *! | | a. [CP] |
| b. [IP] | *! | | b. [IP] |

(Baković 1997: 4)
A couple of criticisms should be mentioned here. One is specific to Bakovic's analysis of the optionality of English complementiser *that*, the other is of a theoretical nature and thus global, and both comments are related to each other. First, we are left to wonder where the MARK-CP constraint is located in the analysis of compulsory *that*, and with parallel concern, how MARK-IP figures into the working of prohibited *that*. Surely, it would not be beneficial to assert that MARK-CP does not exist in the former and MARK-IP in the latter. If this were so, then we would be making an unrealistic and perilous assumption that an unbiased mechanism, such as that proposed under Optimality Theory, would know under what linguistic conditions and with which types of inputs to ignore certain constraints. This seems groundless, and is neither desirable nor fruitful in investigating the situation. We must insist that all three constraints, MARK-IP, MARK-CP, and FAITH[SUB], should always be present in the grammar, and let their strengths of influence be governed by their ranking relative to the other constraints.

We are thus forced to infer that MARK-CP is ranked between MARK-IP and FAITH[SUB] when complementiser *that* is obligatory. It must have priority over FAITH[SUB], or else *that* would be optional, as we explored in (21). Additionally, it must rank below MARK-IP in order to produce the CP candidate as the optimal one in the output. Under the same argumentation, we take it upon ourselves to assume that MARK-IP ranks under MARK-CP but above FAITH[SUB] when the absence of *that* is called for. In summary, to account for the required presence, forced absence, and true optionality of the English complementiser *that*, we resort to three different orderings of the constraints involved, as dictated by the inputs and desired optimal candidate.

This observation leads us directly to our second criticism. It is constructive and justifiable to adopt a framework whereby cross-linguistic variation in a certain linguistic item can be successfully accounted for by distinct rankings of the relevant constraints in the different languages. It is the unique ordering of these constraints that constitute the grammar of a particular language. Because of this, we must keep the ordering of the constraints consistent within any one language, independent of the type of input or the optimal candidate desired. This is crucial if we want to maintain that Optimality Theory is able to account for syntactic considerations. Therefore, we cannot tolerate a situation
whereby more than one ordering of certain constraints exists within a single language in order to explain the observed data.

The ideal solution is to establish however many constraints are needed to account for the phenomenon in question, and then to create one ranking of these constraints, in the correct order. Such a system will be successful if, regardless of the type of input given, the output predicted by this unitary ordering of constraints will always prove to be the optimal one.

Despite the shortcomings mentioned above, we are nevertheless encouraged that there is the potential for true optionality to be properly accounted for within an Optimality Theoretic framework, once the inconsistencies have been addressed. This review of Bakovič's work is pertinent to the present research not only because it highlights the necessity for constraint ranking consistency within a language and constraint ranking variation across languages, but also because, as we shall witness in future chapters, there will be occasions where the presence or absence of a subject pronoun will make no discernible difference in the grammatical judgements of our informants. If we are striving to formulate an Optimality Theoretic framework to account for prohibitions in null subject pronoun usage, we must also be prepared for how to deal with cases where lexical subject pronouns are permitted but not obligatory.

### 3.7 Cole (2000)

We now review Cole (2000), the work upon which the present research is primarily based. Cole, following the example of the more recent works summarised here, also eschews a traditional syntactic framework to account for null subject patterns, in favour of Optimality Theory. He justifies this by alluding to the shortcomings of an overly rigid Government and Binding model, in line with similar arguments already explored earlier in this chapter, as well as the limitations of Minimalism, citing evidence against the ability of the Minimalist Program to account for certain verb-raising and subject pronoun data, as well as for Cole's own hypotheses on morphological maximality, semantic identification, and true optionality (2000: 192ff.).
We first investigate Cole’s proposals on default interpretation, morphological maximality, and semantic identification, highlighting the extent of their tenability. This is followed by Cole’s assessment of Samek-Lodovici (1996), whose Optimality Theoretic methodology Cole largely adopts in his own work, with appropriate modifications. Next, a summary of Accessibility Theory, as constructed by Ariel (1988; 1990), which provides formidable influence on Cole’s approach, is offered. Finally, we present Cole’s empirical research, commenting on its strengths and hindrances, as well as how we choose to build upon his contributions in this dissertation.

3.7.1 Default interpretations

Cole strongly maintains that default interpretations with ambiguous verbs do exist (2000: 100). He presents supposed instances of such default interpretations to help explain observed co-referencing patterns between pro and its possible antecedents. That such default interpretations exist, we claim, is a doubtful conjecture, at least in the scheme by which Cole proposes them.

Cole offers Italian data to exemplify this notion of default interpretation. In the following sentence, the embedded verb sappia is the present subjunctive of all three singular persons, meaning that pro should be able to take as possible antecedents either the third person Maria or the first person singular pronoun io. However, Cole’s stance is that only coreference with io is acceptable:

(24a) Maria ed io1 siamo amici. Bisogna che pro1 sappia la verità.
(24b) *Maria1 ed io siamo amici. Bisogna che pro1 sappia la verità.

‘Mary and I are friends. It is necessary that I/*she know the truth.’

(Cole 2000: 99)
He continues that if co-reference with Mary is intended, an overt third person feminine singular pronoun _lei_ must be expressed:

(24c) _Maria_1 ed _io_ siamo amici. Bisogna che _lei_1 sappia la verità.

‘Mary and I are friends. It is necessary that she know the truth.’

(Cole 2000: 99)

It is not clear under Cole’s analysis why and how the first person is the default subject interpretations, or under what scope. Does this apply just for the present subjunctive verb, or is this practice characteristic of all such Italian constructions where person ambiguity is present? If the first person singular were the accepted default subject of _sappia_, then the presence of _io_ would surely be redundant, so why use it at all? A first person singular overt subject pronoun would violate syntactic and pragmatic principles of economy, especially if no emphasis or contrast was intended, as Cole implies. Additional questions do remain.

Perhaps deictic considerations are in effect. The discourse participants are more salient in the conversation, so in the absence of unambiguous co-reference, we revert by default to the first and second persons. Cole does amend his main constraint, DROPSUBJECT, to include a deictic component (2000: 233). However, Cole’s own comparison between competing first and second person antecedents shows that even the two discourse participants are not treated equally with regard to default interpretations, so deictic factors cannot account for everything here. The example below shows that _sapessi_, an Italian imperfect subjunctive theoretically co-referential with either _io_ or the second person singular pronoun _tu_, takes _tu_ as the default, leaving an overt _io_ to be expressed if it is the intended subject:

(25a) _Tu_1 ed _io_ eravamo amici. Bisognava che pro_1 sapessi la verità.

(25b) *_Tu_ ed _io_1 eravamo amici. Bisognava che pro_1 sapessi la verità.

‘You and I were friends. It was necessary that you/*I know the truth.’
What is most remarkable about the above data is that it directly contradicts traditional approaches, as stipulated by, *inter alia*, Cordin and Calabrese (1988: 540), that Italian requires, under normal circumstances, an overt subject pronoun *tu* in subjunctive clauses. Of course, disagreement is not a problem if the evidence used to disprove previously existing claims is dependable. However, even Cole admits that not all of the Italian native speakers he consulted in his research agreed on the default interpretations. Rather, the claims made above are based on majority opinion, and in fact, one of his informants mentioned explicitly that the example sentences were unclear and would need an overt subject to be interpretable, regardless of the intended antecedent (2000: 100). It is worth mentioning as well that these are clearly contrived sentences, artificially constructed by linguists. For example, it is far more common in standard colloquial Italian to use ‘*bisogna* + infinitive’ than ‘*bisogna*(va) *che*’, the latter normally being replaced by the use of *dovere*. A more natural rendering would have been to use a different main clause predicate, such as *convenire*.

This disharmony is confirmed by informal peripheral work conducted on default interpretations under the auspices of the present research. The informants surveyed did not all agree with these default interpretations in Italian, or, for that matter, in the equivalent Spanish, Portuguese, and Catalan sentences either. This was true both within any one language, as well as in comparison across different languages. Some doubted whether such a determination could even be made. While we do not necessarily doubt that individual speakers may have default interpretations, leaving this for an item of future research, it is highly questionable whether inter-speaker judgements on default references within any particular language are consistent enough to assert effectively for
their existence. Furthermore, we will not necessarily be working with previous idealised or prescriptive definitions of pro usage.

Within Cole’s final framework is a revised version of DROPSUBJECT, which contains an exemption qualification whereby this constraint is fulfilled “except where [fulfilling the constraint] gives an unwanted default interpretation of the unparsed subject” (2000: 233). We maintain that any framework attempting to explain language-internal and cross-linguistic null subject variation that incorporates a component based on default interpretation is unlikely to be successful, unless, of course, it can first adequately account for the empirical discrepancies and theoretical shortcomings mentioned above. As Cole has not resolved these issues in a convincing manner, this can only distract from the success of his overall framework.

3.7.2 Morphological maximality

A crucial component of Cole’s model explaining null thematic subject licensing involves morphological maximality (2000: 121-123). This concept, a formal attempt at summarising Samek-Lodovici’s conclusion in Section 3.4.5 on making full use of available feature morphology, is framed as a discretely increasing scale of the extent to which languages exhibit feature agreement between a verb and its external argument. At the bottom of the gradient is Japanese, where no overt morphological agreement is visible. One step above it is Bengali, a language that shows only person agreement, so verbs are able to indicate whether its subject is first, second, or third person, but there is no indication of gender or number. Cole cites Spanish and Italian as occupying the next highest position on the scale (although we question below the unequivocal classification of Italian in this category), since these languages show person and number agreement between subject and verb. At the apex are Tarifit and Tamazight Berber, languages that display agreement with respect to person, number, and gender. Each language has its own level of morphological maximality, and while this is the attested end of the scale, Cole leaves open the possibility that we may discover languages ranking higher than the

13 Cole spells this as “Tamaziqt” throughout his work. We adhere to the more popular orthography here.
two versions of Berber, showing agreement in person, number, gender, and “features X, Y, Z, ad infinitum” (2000: 122).

Cole proposes that the level of morphological maximality is inversely proportional to the need for context to identify pro. Thus, the placement of Japanese at the lowest level implies that context alone must identify pro. If context is unable to do so, then an overt subject must be used. The exact nature of this context and how it is judged to have successfully completed its identification task are not clear at the moment; we return to this shortly. Bengali will tolerate a null subject if the verb can identify it as a match in the person feature, as this is its level of morphological maximality, but it will need to resort to context to determine whether this potential antecedent satisfies number and gender specifications as well. The Spanish null subject is licensed if verbal desinence can identify, on the basis of person and number agreement, a unique antecedent. This must be accompanied by the role of context to account for the gender feature. Finally, we infer that the two versions of Berber do not need to rely on context at all, as verbal morphology is able to identify person, number, and gender (assuming there are no such hypothetical features X, Y, or Z that Cole alludes to above).

Thus, morphological maximality is a method of formalising the result shown by Gilligan (1987) and others of the non-correlation between richness of verbal agreement morphology and the availability of the null subject. Furthermore, it takes into account that this concept involves gradients, as it accommodates languages showing no overt morphological agreement, those displaying desinences indicating person, number, and gender, as well as those languages that can only do for a subset of those three features.

Returning to the issue of null thematic subject licensing, Cole contends that “a unified solution involving both agreement and context is needed, since neither on its own is sufficient to account for the data” (2000: 117). He formulates his two-part proposal as follows:
Subject thematic *pro* may occur when:

(a) its identification is completed by AGR, up to the point of morphological maximality (with a limited tolerance for syncretisms); and

(b) it is semantically identified by the features of an accessible antecedent.

Provided that the entity identified at (a) is the same as that identified at (b)

(Cole 2000: 172)

It should be pointed out that, with regard to the first stipulation, Cole is not consistent throughout his entire work. Sometimes his definition will appear as above; at other instances, he will delete references to AGR, without any intended distinction observed. It is not certain whether this is merely typographical oversight or if it is deliberate. We mention this point because it poses non-trivial ramifications and provides a potential point of exploration for further research. Depending on which definition is used, we may claim that the Romance languages we examine in this work show two different levels of morphological maximality. We have Spanish and Portuguese, whose level of morphological maximality is person and number. In addition, we investigate two languages that may be one degree higher: Catalan and Italian, which permit the expression of gender in the verbal construction, the former by past participle agreement with a preceding direct object in perfective tenses, the latter being able to do so in that manner, as well as in perfective unaccusative structures. It is in this respect that Cole’s toggling mention of AGR, is significant. Its inclusion rules out classifying Catalan and Italian in a category above Spanish and Portuguese. By excluding it, however, it leaves open the possibility that Catalan and Italian may be at least partially classified in the same category as Berber, as long as we consider verbal adjectives like past participles separately from pure adjectives, which are generally unable to indicate agreement with respect to all three features. We hope to identify both the extent of the role of morphological maximality in null subject licensing, as well as the validity of Cole’s framework as a whole.
We now have a better idea of what is intended in the first constituent of Cole's proposal on null thematic licensing. As for the second part of his formulation, it remains to be explored what is meant by the terms "semantically identified" and "accessible antecedent". We turn to these two items in the following subsections.

3.7.3 Semantic identification

There are three principal doubts concerning Cole's characterisation of the role of semantic identification in null subject pronoun licensing that are left unresolved in his work. One problem is related to inconsistencies within his explanation; the other two issues arise as a result of a lack of the same.

We first address some pitfalls in the data Cole provides. He defends the ungrammaticality of (27) below by pointing out that the number specification between the plural subject of the first sentence and the singular-feature bearing pro in the second sentence:

(27)  *[Juan y Juanita] llegaron a casa. Pro_1 abrió la puerta.

'Juan and Juanita arrived at home. (pro) opened the door.'

(Cole 2000: 125)

While the ungrammaticality of the sentence, at least as judged by our Castilian-speaking informants, is not being contested, the forced co-referencing contained within it is. Cole does not explain why pro is unable to access either individual component within the compound subject, and is coerced into feature-matching with the two individuals taken as one third person plural unit. Can we claim here that no suitable antecedent is available here for co-referencing?

Under similar logic, the following acceptable sentence (28a) should prove ungrammatical, as shown in (28b) if similarly forced co-referencing took place:
(28a)   *Juan y yo llegamos a casa. Pro abrí la puerta.
(28b)   *[Juan y yo]₁ llegamos a casa. Pro₁ abrí la puerta.

‘Juan and I arrived at home. (I) opened the door.’

The grammaticality of (28a) obtains because pro is indeed able to access one single member of the first person plural subject. First person singular yo is a suitable coreferent available to act as an antecedent to pro at the beginning of the second sentence. Resolution of this discrepancy would lend greater support to Cole’s proposal on null subject licensing.

The examples above also relate to the overlap between the functions performed by morphological maximality and semantic identification. The feature in question in (27) is number; Cole claims that the sentence is ungrammatical because of a mismatch between the plural subject in the first sentence and the singular one in the second. According to the formulation of Cole’s hypothesis, however, semantic identification should only come into effect after morphological maximality has done its part. The person feature is within the scope of morphological maximality in Spanish, so we are left to wonder why semantic identification is even involved in details on person. As currently proposed by Cole, it should only be handling the gender feature, since this falls out of bounds of morphological maximality in Spanish. This questions whether morphological maximality and semantic identification are mutually exclusive realms. A perhaps weaker, but surely more accurate revision, might be to amend the theoretical background justifying his proposal by requiring that semantic identification act not solely on the features not covered by a language’s morphological maximality, but on all three features. We leave this debate open for the time being.

The second issue concerns the differences in null subject licensing between two languages at the same level of morphological maximality. Cole notes that in Japanese, an antecedent in context can semantically identify pro, whereas in Norwegian, this is not possible (2000: 129). Japanese and Norwegian, both belonging to the zero morphological
maximality category, must therefore diverge on the second component of Cole’s licensing proposal in order to account for their null subject differences. What is not explored, however, is the nature of this difference. Is it simply parametric, that Japanese possesses a positively set parameter with regard to the ability to identify pro through context alone, whereas Norwegian has the negative specification on this parameter? If so, then does this fundamentally contravene one of the founding motivations of our area of research, that is, the move away from parameterising these cross-linguistic differences in the first place? In other words, have we just substituted one parameter for another?

There is no treatment of this issue whatsoever. If this account is indeed what Cole implies, then he offers no independent justification for such settings. We ponder why it is not the Scandinavian language with the positive setting and the Asian language with the negative one.

Finally, we are left to investigate the nature of semantic identification. There is no discussion on whether all types of context operate in the same manner. Does identification by gender, within a more liberal interpretation of verbal morphology, such as Italian past participle agreement and perfective unaccusatives, work the same way as identification in a gender-inherent predicate of a clause, for example, being pregnant or becoming a mother? The question arises as well as to whether direct identification of a possible antecedent works in a parallel manner as identification by default or deduction, in other words, the elimination of all other possible antecedents? These are uncertainties that Cole does not treat directly, and attempts will be made in the present research to formalise an approach incorporating these questions.

3.7.4 Ariel (1988; 1990) and Accessibility Theory

We now turn to the other crucial component in Cole’s second condition for null subject pronoun licensing, the idea of accessible antecedents. Cole’s definition of accessibility is based on the theory of the same name, proposed by Ariel (1988; 1990). It consists of four factors:
i) Distance: The distance between the antecedent and the anaphor

ii) Competition: The number of competitors on the role of antecedent

iii) Saliency: The antecedent being a salient referent, mainly whether it is a topic or non-topic

iv) Unity: The antecedent being within vs. without the same frame/world/point of view/segment or paragraph as the anaphor

(Ariel 1990: 28-29)

Each of these four determinants may display a range of accessibility levels. For example, the shorter the distance between a subject and its intended antecedent, the more accessible we consider the antecedent to be. This seems intuitive enough; the more proximate an antecedent is, be this temporally or spatially, the easier it is to be retrieved when attempting to resolve which preceding entity a subject is intended to be co-referenced with.

Competition appears to be a relevant factor to accessibility as well. It is a reasonable argument that the fewer the number of available candidates for co-referencing there are, the easier the task will be for subjects to identify their antecedents, and hence the more accessible these will be. A subject will be able to identify its antecedent, more quickly and with less effort, if only a single one is available, than will another subject with multiple potential antecedents, where more work is required to determine which of the candidates is the most suitable.

As for saliency, topical antecedents are considered to be more accessible than non-topical ones. This is related to discourse prominence; topics are more focused in utterances than non-topics, and so a subject will be more likely to choose a more conspicuous antecedent than a less noticeable one, all other things being equal.

The concept of unity is harder to gauge because Ariel does not address this component as adequately as the other three. Under one interpretation of Ariel’s fourth element, this involves substantial discourse analysis. In particular, it takes into account how far a potential antecedent can be in a conversation and still act as an available coreference for some following subject, which may occur a number of sentences after it. The unity factor, then, would not be fully independent from the distance element, unless this latter one is interpreted as applying only intra-sententially and not paragraphically.
Another alternative would be to take a loose approach to the "same frame/world/point of view" stipulation by understanding it to refer to our knowledge of reality. That is, we know that in our world, only women may have babies and so null subjects in statements about pregnancy should look for feminine antecedents more readily than for masculine ones, thus making the former more accessible. This condition is clearly not as concrete as the other three and is open to disparate interpretations.

Levels of accessibility having been defined, we now turn to their interaction with the various referring NPs. Ariel suggests a ranking of NPs with regard to accessibility strength (1988: 84). Null pronouns, at the top of this scale, may be co-referenced only with antecedents of highest accessibility, as determined by the four-part classification mentioned previously. Below null pronouns are their overt counterparts, which may take less accessible antecedents for co-reference. At the bottom of the order are full proper name NPs, which normally take antecedents judged to be lowly accessible, as, for example, entities not yet introduced into the discourse.

Ariel provides empirical evidence in the form of quantitative textual analysis in support of her classification scheme (1988: 69ff.). She cites as further evidence the compatibility of her accessibility strength ranking with Givon's (1983: 10) hierarchy of topic continuity and accessibility, which places zero anaphora above unstressed pronouns which, in turn, are positioned above various categories of NPs (Ariel 1990: 74). It is noted as well that there is, generally speaking, an inversely proportional relationship between phonetic content and accessibility strength. Zero pronouns display no phonological material and refer only to the most accessible antecedents, whereas pronouns carry more phonetic substance and are able to refer to less accessible antecedents. Ariel claims that this correlation holds because phonologically weak units, such as null and subject pronouns, are able to identify highly accessible antecedents and hence do not need to contain much lexical substance for successful retrieval (1988: 82). Whether this piece of independent justification for her hierarchy is convincing is debatable.

We are now in a position to evaluate Cole's flow chart guiding overt thematic subject pronoun licensing, as proposed below:
thematic subject \textit{pro} is identified up to the point of morphological maximality or as far as possible towards it by AGRs

identification as at (i) is supplemented by semantic identification from an antecedent in context

where a combination of (i) and (ii) fail to identify thematic subject \textit{pro}, it may have a default interpretation (as in Italian)

where neither (i) and (ii) nor (iii) provide an identity for thematic subject \textit{pro}, it may not occur and an overt pronoun must be substituted

(Cole 2000: 131)

We maintain that this formulation is in need of further refinement. Our hesitations against default interpretation as stipulated in (iii) have already been mentioned. We understand “semantic identification from an antecedent in context” in (ii) to be governed in large part by Ariel’s link between NP type and accessibility strength, but it is still not entirely clear how to account for differences between languages at the same level of morphological maximality.

It is worth returning briefly to what Cole characterises as the primary separation between \textit{pro}-drop and non \textit{pro}-drop languages, their relative contextual strength (2000: 19). Cole mentions that languages that permit \textit{pro} exceed a certain level of contextual strength, this being determined by Ariel’s accessibility hierarchy, whereas those languages that do not are classified as contextually weak (2000: 131-132). This goes back to what we alluded to at the end of Section 3.7.3. While exploration of Ariel’s hierarchy now leads us to conclude that contextual strength is constructed not as a parameter but more like a gradient or cline, it still does not answer, however, how languages are assigned these levels of contextual strength. Are we to infer that null and subject pronouns differ cross-linguistically, so that Japanese NPs are stronger in accessing antecedents than their Norwegian equivalents? Ariel does mention that the “accessibility scale is, to some extent, language-specific” (1998: 83). If this is what is intended, is there independent evidence in support of this cross-linguistic variation? One
could not argue, of course, that this difference in strength is due to the tolerance in the Japanese language for null subjects and the opposite in Norwegian, because this reasoning would be circular. We are not sure, then, where the argumentation in support of differing contextual strength within various languages begins and ends.

In referring back to Sigurðsson’s presentation on the loss of pro-drop in Icelandic in Section 3.3, we saw that the prohibition on null thematic subjects in modern Icelandic arose as a result of the loss of the ability to license pro through free indexing of an NP in preceding discourse. Cole asserts that this could be interpreted as Icelandic changing diachronically from a contextually strong language to a contextually weak one (2000: 132). Do we then assume that the basis by which contextual strength is defined is dependent on the method by which pro is licensed?

Another question does arise, and is terminological in nature. Ariel does employ the two terms “accessibility” and “saliency” fairly interchangeably (e.g., 1998: 80), and we note the danger of this usage when saliency is one of the components defining accessibility. We recall here our investigation in Section 2.3.3 that saliency does not have a direct influence on the licensing of pro-drop. Thus, it is a natural extension to see whether the synonymous usage of these two words is tenable. If the two are indeed largely equivalent, then we wonder whether adopting Ariel’s Accessibility Theory in a framework explaining null subjecthood is consistent and resolvable with our discoveries on saliency in the previous chapter.

The present research hopes to clarify whether accessibility can be explained by distance, competition, salience, and unity, as defined by Ariel. We also attempt to discover how well this accessibility theory can be successfully incorporated into an account of null subject licensing.

3.7.5 Cole’s comments on Samek-Lodovici (1996)

We return our focus now to the Optimality Theoretic framework used to describe null subject availability. As Cole’s work on null subject licensing is primarily based on the model offered in Samek-Lodovici (1996), we first summarise Cole’s comments on
the strengths and weaknesses of Samek-Lodovici's proposal. This sets up a review of the ways Cole chooses to modify Samek-Lodovici’s framework in order to present what he considers to be an improved version.

Samek-Lodovici (1996: 89) acknowledges that there is considerable overlap between subjecthood and topichood, pointing to data supplied by Calabrese (1985) on coreferencing patterns in addition to Saccon's (1993) suggestion of the intrinsic nature of preverbal subjects to be topics. However, he correctly distinguishes that the two are not strictly synonymous, citing discrepancies in Calabrese's explanation to support topichood and not subjecthood as the pertinent factor in tolerating null subjects. Samek-Lodovici shows that in the following example, Calabrese's hypothesis would predict coreference between the null subject and the subject of the adjunct clause Sandro, which is the theme, whereas grammaticality actually occurs only when the null subject takes the non-theme topic in the left-dislocated phrase Mario as the antecedent:

(30) Mario, quando Sandro, l'ha incontrato, c; / ??c; è arrossito.

Mario, when Sandro him-has met, (he) is blushed

'As for Mario, when Sandro met him, (he = Mario/ ??he = Sandro) blushed.'

(Samek-Lodovici 1996: 92-93)

Based on this and other arguments, Samek-Lodovici adopts a model centring on topics and not subjects. Cole, on the other hand, supports using subjecthood and not topichood as the leading factor determining null subject licensing. He appeals to parallel texts in Italian and Spanish to show that where co-reference as proposed by Samek-Lodovici's topic-dominant hypothesis should occur with the most recent topic, it is in fact the most recent subject that serves as the appropriate antecedent (Cole 2000: 223-225). This leads Cole to replace Samek-Lodovici's crucial constraint DROPTOPIC in favour of his own DROPSUBJECT, which calls for the non-realisation of an "unfocused pronominal subject co-referring with the last recent sentence topic that provides a semantically adequate subject to the verb of the current sentence" (2000: 226). We
explore the possibility that this constraint is too complex to be tenable theoretically, and that it can be significantly improved upon through appropriate simplification, either within itself or by being broken down into multiple constraints. What is more worrying about DROPSUBJECT is its mention of the concept of topic within the definition itself. This weakens Cole’s justification of preferring his constraint over Samek-Lodovici’s DROPTOPIC, as the principle benefit that Cole asserts in DROPSUBJECT over DROPTOPIC is that subjectionhood and not topichood is more directly responsible for null subject licensing. We return to this debate between DROPSUBJECT and DROPTOPIC in Chapter 6.

In addition to DROPSUBJECT, Cole offers five additional variations of it to complete the repertoire of constraints relevant to null subject licensing. Two of them do not appear to be germane to our investigation. DROPSUBJECT2 is fulfilled when the null subject is co-referenced with a preceding direct or indirect object (Cole 2000: 229), and DROPSUBJECT4 is obeyed when the same occurs with the agent of a previous passive sentence (Cole 2000: 243). The first hesitation is that these constraints are rather ad hoc and not theoretically rigorous enough to contribute to a fundamental understanding of null subjectionhood. We seek a more concise and sharper formulation of these constraints. The second concern involves what we have already seen in Chapter 2, that the grammatical relation of the antecedent in the previous sentence does not have a direct bearing on its ability to serve as a co-referent to a subsequent null subject, at least with regard to the non-subject positions direct object, indirect object, and oblique. We do acknowledge that, within some languages, a distinction may be made between subject and non-subject antecedents, as will be explored in the following chapters.

There are three variations of DROPSUBJECT that may be useful to our Optimality Theoretic framework. DROPSUBJECT3 is appeased when the null subject is identified up to morphological maximality by unique agreement (Cole 2000: 240). DROPSUBJECT5 is complied with when, by default, the possessor of the topic of the previous sentence provides its only possible antecedent (Cole 2000: 244). Finally, DROPSUBJECT6 is similar to DROPSUBJECT5, except it is the possessor of the topic, and not the topic itself, that is the only possible antecedent by default (Cole 2000: 245).
These three constraints, while leaving a bit to be desired regarding their wording and theoretical approaches, do show some potential in our Optimality Theoretic model. We continue with them for the moment, adjusting for appropriate modifications, in our investigation of null subject licensing. The relevance of these constraints will be considered in greater depth in Chapter 6, when we provide a more complete exposition of how they figure in our proposed framework. They do appear to be more directly pertinent to our study than DROPSUBJECT2 and DROPSUBJECT4, and for this reason we no longer continue to consider these two constraints seriously in our framework.

Cole takes Baković's lead in accounting for true optionality in null subject licensing by adopting a solution parallel to that provided to explain the permitted, but not required, usage of English complementiser that. He first considers ranking DROPSUBJECT and PARSE on equal footing (Cole 2000: 234). We recall Samek-Lodovici's demonstration that the crucial factor licensing null subjects was the prioritised ranking of DROPTOPIC (the predecessor of DROPSUBJECT) over PARSE. That Italian permitted null subjects was accounted for by DROPTOPIC >> PARSE, whereas English could not because it ranked these two constraints in the opposite order. The result of equalising DROPSUBJECT and PARSE is that this is not inconsistent with either directional ranking and thus gives the appearance of true optionality, allowing for both null subjects as in Italian and their overt counterparts as in English. However, postulating DROPSUBJECT and PARSE as tied constraints would then leave the grammar unable to account for situations where the subject pronoun must or must not be used. In effect, resolving the optionality problem by proposing tied constraints would come at the expense of undoing all the progress achieved in predicting null subject usage in the first place.

Cole therefore concludes that Bakovic's solution is more appealing. He devises a constraint that functions identically to Bakovic's FAITH[SUB] and calls it FAITHFULNESS(DEFPRO), whose settings depend on whether the "subject pronoun whose omission would give rise to an identical interpretation by default as its realisation does" appears in the input or not (Cole 2000: 238). It is fulfilled if the input setting matches that in the output. One hazard in this approach is that this opens up the possibility of there being more than one unique ranking of constraints in a language,
which is incompatible with the central tenets of Optimality Theory\textsuperscript{14}. We examine in due course both the necessity of applying Bakovic's solution on true optionality to our framework, as well as the relevance of Cole's constraint FAITHFULNESS(DEFPRO) itself.

\textbf{3.7.6 Cole's question types}

Our review of Cole (2000) concludes with a quick survey of his methodology. Cole gathers his data by testing co-referencing possibilities in thirteen different syntactic constructions for each of the following languages: Spanish, Italian, Greek, Hebrew, Serbo-Croat, Chichewa, Chitonga, and Chitumbuka. These constructions are outlined below; the subject pronoun in parentheses is the item being tested for covertness:

\textbf{Type A: The subject of the second sentence is co-referent with a topic in the first.}

Example: Every morning Gianni\textsubscript{1} visits the museum. In the afternoon (he\textsubscript{1}) visits the university.

\textbf{Type B: The topic of the second sentence is co-referent with the agent of the passive in the first.}

Example: Every morning the museum is visited by the president\textsubscript{1}. In the afternoon (he\textsubscript{1}) visits the university.

\textbf{Type C: The subject of the second sentence is co-referent with the topic of the first which has a possessor.}

Example: John\textsubscript{2}'s father\textsubscript{1} is terrible. (He\textsubscript{1}) hates him\textsubscript{2}.

\textsuperscript{14} Note that having multiple grammars in a language does not imply having one ranking of constraints for each of those grammars. Multiple grammars, such as the optionality available in English with \textit{that}-deletion or preposition stranding, should still be accounted for by one unique, language-specific ranking of constraints in Optimality Theory. It is the interaction of the fixed constraints, and the manner in which those constraints are formulated, that result in optionality, not the existence of multiple sets of constraint rankings. See McCarthy (2002: 230) for a list of various treatments of this optionality conundrum in Optimality Theory.
Type D: The subject of the second sentence is coreferential with the possessor of the topic in the first, with the object clitic or AGR₀ of the second sentence coreferent with the topic of the first sentence and capable of differentiating between that topic and its possessor.

Example: \(\text{John}_1\)'s mother\(_2\) is terrible. \((\text{He}_1)\) hates her\(_2\).

Type E: Like D but with the object clitic or AGR₀ in the second sentence incapable of differentiating between the topic of the first sentence and its possessor.

Example: \(\text{John}_1\)'s father\(_2\) is terrible. \((\text{He}_1)\) hates him\(_2\).

Type F: The subject of the subordinate clause is co-referent with a topic which has a possessor.

Example: \(\text{John}_2\)'s father\(_1\), we know the reason why \((\text{he}_1)\) hates him\(_2\).

Type G: The subject of the subordinate clause is co-referent with the possessor of the topic, and the object clitic or AGR₀ in the subordinate clause is capable of differentiating between the topic and its possessor and corefers with that topic.

Example: \(\text{John}_1\)'s mother\(_2\), we know the reason why \((\text{he}_1)\) hates her\(_2\).

Type H: Like G, but with the clitic of AGR₀ in the subordinate clause incapable of differentiating between the topic and its possessor.

Example: \(\text{John}_1\)'s father\(_2\), we know the reason why \((\text{he}_1)\) hates him\(_2\).

Type I: The subject is co-referent with a dislocated topic that has a possessor

Example: \(\text{John}_2\)'s father\(_1\), \((\text{he}_1)\) hates him\(_2\).

Type J: The subject is co-referent with the possessor of the topic, and an object clitic or AGR₀ is coreferent with the topic and capable of differentiating between it and its possessor

Example: \(\text{John}_1\)'s mother\(_2\), \((\text{he}_1)\) hates her\(_2\).
Type K: Like H, but with the object clitic of AGR₀ incapable of differentiating between the topic and the possessor

Example: \( \text{John}_1 \)'s father₂, (he₁) hates him₂.

Type L\(^{15} \): A question has a subject/topic that has a possessor. The answer has an object clitic capable of differentiating between the topic and the possessor, but actually co-referent with the subject/topic of the question, and its subject is co-referent with the possessor of that topic.

Example: Question: Has John₁'s mother₂ left?
Answer: No, (he₁) invites her₂ to dinner on Sunday.

Type M: A question with an answer, as L, but the AGR₀ or object clitic in answer is incapable of differentiating between the subject of the question and its possessor.

Example: Question: Has John₁'s father₂ left?
Answer: No, (he₁/₂) invites him₂/₁ to dinner on Sunday.

(Cole 2000: 134-136)

A concise recap here summarising the typology of some of the constructions above would be helpful in a clearer understanding of this seeming disarray of syntactic structures. We note first the minimal pairs that exist. With regard to gender, these are D versus E, G versus H, and J versus K. There are no counterparts for C, F, and I for this feature. Alternations in topic and non-topic antecedent only are present in C versus E, F versus H, and I versus K. D, G, and J do not have analogous types that differ in just the choice of antecedent alone. Finally, minimal groups with respect to distance between subject and co-referent are as follows: [C, F, I], [D, G, J], and [E, K, H]. These observations are laid out in table format below:

\(^{15}\) Cole's Type L comes in two versions. Type L(i) is an instance of same reference, where the topic of the question appears as the subject of the answer. Type L(ii) shows switch reference, where the answer in the subject is actually the possessor of the topic in the question. What we refer to throughout the text as Type L is, more specifically, Cole's Type L(ii).
(31)

<table>
<thead>
<tr>
<th>ANTECEDENT</th>
<th>GENDER</th>
<th>2S</th>
<th>T, MC, SC</th>
<th>T, MC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject co-refers with topic</td>
<td>Object clitic can</td>
<td>(27)</td>
<td>(19)</td>
<td>(3)</td>
</tr>
<tr>
<td></td>
<td>differentiate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objective clitic cannot</td>
<td>C</td>
<td>F</td>
<td>I</td>
<td></td>
</tr>
<tr>
<td>differentiate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject co-refers with</td>
<td>Object clitic can</td>
<td>D</td>
<td>G</td>
<td>J</td>
</tr>
<tr>
<td>possessor</td>
<td>differentiate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>H</td>
<td>K</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Object clitic cannot</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>differentiate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

COLUMN KEY:

I. ANTECEDENT: whether the subject co-refers to the topic or its possessor

II. GENDER: whether the clitic can differentiate between the topic and its possessor on the basis of gender

III. 2S: a construction with two single clause sentences

IV. T, MC, SC: a construction with a left dislocated topic, a main clause, and a subordinate clause

V. T, MC: a construction with a left dislocated topic followed by main clause

The three gaps left by Cole's methodology in the chart are filled in above by numbers in parentheses. We use these numbers in the next chapter to refer back to the corresponding spaces in this chart. It should be pointed out as well that L and M differ only by gender of the topic, and that M acts as a minimal pair with itself with respect to topichood of antecedent, depending on the choice of indexing selected. They do not appear in the table because it is clear that Types L and M are inherently different from those types that are included above. Our second observation is that the only difference between Type A and Type B is that the second sentence of the latter is the passivised version of the same in the first.

We discuss here the four syntactic constructions, Types A, B, L, and M, that we do not employ in the methodology of the present research, along with our reasons for omitting them. Descriptions of the types that we do use will be given when we present
our corresponding original data in the next chapter. Type A appears to be the default case testing for the availability of null subjecthood. We maintain that languages permitting a null thematic subject in Type A are worthy of inclusion in any pro-drop study, and those that do not allow it are safely classified as non-NSLs. All eight of the languages Cole uses in his study tolerate pro in Type A, and all six Romance varieties that we resort to in our research allow for the null subject there as well. This leads us to conclude that Type A is not sufficiently discriminating for our purposes to be of any diagnostic utility.

Type B is concerned with whether the co-referent of the second sentence subject, now presented not as the subject of the first sentence but as an agent in a by-clause passive, can still license pro. The reader should be convinced by now that the results obtained from this type are unlikely to prove informative to our research. We have already explored the lack of direct influence of grammatical relations and agency in the previous and present chapters.

According to Cole’s Type B data from the eight languages, only Italian prohibits the use of pro as subject of the second sentence. We argue that this has less to do with the status of the antecedent being an agent in a by-clause passive in the first sentence, than with the fact that in Italian, mostra, exhibition, contains feminine singular third person features, and thus shows overlap between the masculine singular third person Gianni. The Italian informants consulted in the present research do not agree with Cole on the grammaticality of (32a). Some claim that it is in fact acceptable with pro. Others assert that (32a) may be a bit awkward and marginal without overt lui, but they would not go so far as to say that it is definitely ungrammatical.

The usage of an overt subject pronoun, then, is based not on the agentive nature of the antecedent Gianni, which is supposedly the principal item under consideration in Type B, but rather on the coincidental matching of certain features of the two possible antecedents in the first sentence (in this case, person and number). But we are already testing for this in Type D. If we change “exhibition” to “churches”, which is feminine plural in Italian, pro becomes available as there is no possible competition between the two visible antecedents, visita indicating that a singular subject is intended:
(32a) Ogni mattina la mostra è visitata da Gianni. Nel pomeriggio *pro1/lui, visita l’università.

(Cole 2000: 144)

(32b) Ogni mattina le chiese sono visitate da Gianni. Nel pomeriggio (pro1) visita l’università.

All of our native speakers concur that (32b) is grammatical and clearly more acceptable than (32a).

It is not clear what the purposes of Types L and M are. Samek-Lodovici provides data from question and answer pairs in his justification that the subject in the answer, being the counterpart to the wh-phrase in the question, is a contrastively focused element (1996: 113). This assertion is important to Samek-Lodovici’s defence of his model. However, Cole does not posit questions with wh-phrases in his Types L and M. The answers in these two types are of the yes/no variety, so they do not deal with contrastive focus. On the surface, then, there appears to be no substantial discrepancy between Types L and M, and Types C, D, and E, which are also two-sentence constructions testing possible co-reference with (un)ambiguous antecedents in the presence of (in)distinguishing clitics. There are no differences, in any of his eight languages, between Cole’s results collected from Types L and M, and those taken from Types C, D, and E. As a result, we do not expect any significant contribution to result from Types L and M that will not already appear through investigation of Types C, D, and E. For this reason, we do not consider Types L and M any further.

3.8 The present research

A treatment of the methodological specifics of this work begins here. We first draw attention to what we keep from Cole (2000), along with which aspects of his research differs from ours. Justifications for such departures from Cole’s procedures
follow. We end the chapter with a discussion of how the data for this research were collected. This sets up the presentation of our data in Chapter 4.

3.8.1 How it differs from previous studies

The most noteworthy distinction between the present research and Cole (2000) is that we expand on the range of his language-internal data. We saw in the previous subsection that the only differentiating factor that Cole uses between any two competing antecedents is gender. Our approach is to generalise and expand this treatment to the other two features as well. We thus investigate the effects of person and number in parallel manner, in addition to various combinations of the three features. The motivation for such an attempt is that we can test in a systematic format Gilligan’s and Samek-Lodovici’s related predictions on the relative strengths of person, number, and gender features. We also pay close attention to the manner in which these features are marked, keeping in mind that they may appear through clitic choice or inflectional desinence, which may not operate in the same way. Finally, it is believed that the information gathered from our data will be particularly informative whenever syncretic verb forms are available, as this will give us further insight on how verbal ambiguity is resolved.

The linguistic scope taken in this work is pan-Romance. We choose this practice for two reasons. The first is that by looking at closely related languages, we are able to pinpoint the existing microvariation that would otherwise not be observable in a comparison of more distant null subject languages. This is our logic in narrowing down the range of languages from Cole’s assortment of Indo-European and African languages to just Romance. Second, we have chosen this particular group of languages in acknowledgement of the long tradition, popularity, and accessibility of Romance philology. Much has been done on the descendant languages of Latin, so we realise that while there has been much uncovered, there is also plenty of room for additional findings. There is a wealth of pertinent information available, not only through the data and results in the existing literature, but also in the native speakers available for our task. It is
asserted that the number of informants willing to provide extensive grammatical judgements, and the satisfactory range of Romance varieties that they are fluent in, are such as to facilitate a project of this magnitude and scope.

This is not to say that our methods could not be applied to non-Romance varieties. Rather, we hope to justify that the conclusions reached on the basis of an examination of our chosen languages can be insightful not only to the Romance varieties we do not test, but also to languages in other families as well. Furthermore, there is no reason to discount the possibility that our research will prove relevant in a comparison across different language families. So we foresee both cross-linguistic and cross-familial typologies as potential outcomes of the conclusions presented at the end of this work.

We have already discussed which of Cole's structure types we keep and which we disregard, with our reasons for this move. To supplement the gaps in Cole's data, we are compelled to create other types to test areas not dealt with by Cole, either through direct omission or imperfect assessment. We point out in particular the underdeveloped research on gender-inherent predicates and null subject licensing as one of the innovative approaches in the present work. A full explanation of this appears in the following chapter.

Finally, the most significant departure from the approach taken by both Samek-Lodovici and Cole lies in our empirical procedure. These two investigate under which conditions null subjects are necessary (e.g., Samek-Lodovici 1996: 73). We argue that less definite answers are given when we ask when null subjects are necessary, because there is wide disagreement on when an overt subject is truly optional, and when it is permitted but considered redundant, awkward, or marked (e.g., Cole 2000: 237). The important distinction to be made is that, while using overt subjects in these situations may not lead to the most natural or optimally graceful expression, their inclusion is, strictly speaking, not ungrammatical. Thus, we are forced to contend with subjective stylistics, which is difficult to account for impartially, and this complicates the question further than is helpful or desirable in our investigation.

We maintain that it is more telling if we examine when null subjects are strictly prohibited, in other words, when lexical subjects are obligatory. Grammatical judgements offered using this approach are normally given less hesitantly and with
greater inter-personal consensus, resulting in more accurate and less controversial data. We thus adopt this as our approach.

3.8.2 Methodology

The data gathered for this work were obtained for six Romance varieties, and for each of them, we had at least two native speakers available to give grammatical judgements. Our primary informants were as follows:

Brazilian Portuguese: (1) man in his mid-20s from São Paulo  
(2) woman in her late-30s from São Paulo  
(3) woman in her late-20s from Belo Horizonte

Castilian:  
(1) woman in her mid-20s from San Sebastián  
(2) man in his mid-20s from Madrid

Catalan:  
(1) woman in her mid-20s from Girona  
(2) woman in her late-20s from Barcelona

European Portuguese: (1) woman in her mid-20s from Lisbon  
(2) man in his mid-20s from Lisbon  
(3) woman in her mid-30s from Oporto

Italian:  
(1) woman in her mid-30s from Rome  
(2) man in his late-20s from Milan

Puerto Rican Spanish: (1) woman in her late-20s from San Juan  
(2) woman in her mid-20s from San Juan

It is worth keeping in mind when considering the distribution of our informants that our research is not one focused on sociolinguistics. Thus, we are not overly concerned with obtaining a representative cross-sample of the population with respect to gender, age, or region.

Each test sentence, lacking an overt subject pronoun in the relevant syntactic position, was evaluated according to the following procedure:
(1) Is the sentence grammatical?

(1a) If not, would including a subject pronoun render it grammatical?

   (1a.i) If so, we conclude that a null subject is prohibited here.
           NOTE RESULT. END

   (1a.ii) If not, we conclude that the sentence is flawed, disregard it,
           and devise an improved one to test the same thing.
           RETURN TO (1).

(1b) If so, GO TO (2).

(2) Can the null subject co-refer to the topic antecedent?

(2a) If so, we conclude co-reference between pro and non-topic is grammatical.
        NOTE RESULT. GO TO (3).

(2b) If not, would including an overt subject pronoun permit the co-reference?

   (2b.i) If so, we conclude that an overt subject pronoun is required for co-
          reference between topic antecedent and subject. NOTE RESULT.
          GO TO (3).

   (2b.ii) If not, we conclude that no co-reference is possible between topic
          antecedent and subject. NOTE RESULT. GO TO (3).

(3) Can the null subject co-refer with the non-topic antecedent?

(3a) If so, we conclude that co-reference between pro and the non-topic
     antecedent is grammatical. NOTE RESULT. END

(3b) If not, would including an overt subject pronoun permit the co-reference?

   (3b.i) If so, we conclude that an overt subject pronoun is required for co-
         reference between non-topic antecedent and subject. NOTE
         RESULT. END.

   (3b.ii) If not, we conclude that no co-reference is possible between non-
          topic antecedent and subject. NOTE RESULT. END.
We provide here a Castilian example as a concrete illustration of the above, using the equivalent of the sentence “John’s father, (he) hates him.”:

(1) Is the sentence grammatical?

*El padre de Juan, (pro) lo odia.

Response: Yes (1b). GO TO (2).

(2) Can the null subject co-refer to the topic antecedent?

*El padre₁ de Juan₁, (pro₁) lo odia.

Response: Yes. We note the result that co-reference between pro and the topic antecedent is grammatical. GO TO (3).

(3) Can the null subject co-refer to the non-topic antecedent?

*El padre₂ de Juan₂, (pro₂) lo odia.

Response: No (3b).

(3b) Would including an overt subject pronoun permit the co-reference?

*El padre de Juan₁, el₁ lo odia.

Response: Yes (3b.i). We note the result that an overt subject pronoun is required for co-reference between an overt subject pronoun and a non-topic antecedent. END.

Therefore, the appropriate evaluation for this item would be that co-reference between the null subject is possible with the topic antecedent, but not with the non-topic antecedent.

Where the judgments given by the two informants did not coincide, a third native speaker of the same Romance variety was approached, after which the two original informants were consulted again for further clarification of their responses. It was
normally the case that general agreement could be reached at this point among the three speakers. In the rare case where the three grammatical judgments could not be reconciled, the majority response was the one accepted, and we were always able to locate a fourth informant to confirm the validity of this action.

This concludes our review of the literature and the explanation of the methodology taken in the present work. We now divert our attention to the original contributions of the present research.
Chapter 4: Presentation of the data and initial observations

We now introduce the data collected in support of this work. The format chosen here is to display the various types by their relative strength of discrimination, in increasing order, such that we begin the chapter examining the types that provide the least contrasting information among the languages. We then progressively introduce more distinguishing types, leading towards a survey of the types that are best able to differentiate between our selected languages at the end of the data presentation. Therefore, the order of types contained here will not be the same as that in Cole (2000). We choose this approach in an attempt to elucidate the specific differences between the languages. Our belief is that this effort will lead to a better appreciation of the language hierarchy that is observed at the end of the chapter.

Within each one of the types, we offer the data in a systematic fashion, starting with sentences that show no feature differences between the topic antecedent and the non-topic antecedent. Next, we use gender as the only feature separating the topic antecedent from the non-topic antecedent. These are the two cases that Cole examines in each of his types. Our improvements on his methodology follow, presenting versions of those sentences using only number to distinguish the two antecedents, and finally, only person. At the end of Section 4.1, we also examine the simultaneous occurrence of more than one of these three features, taken in various combinations. An explanation of why we do not pursue the same for the other types is provided there.

Once the data have been presented, we summarise the patterns observed therein and make some preliminary remarks that will be helpful in a consideration of our theoretical analyses in subsequent chapters.
4.1 Topic followed by Main Clause

These sentences follow the format:

(a) [antecedent2]'s [antecedent1], [antecedent1] hates [antecedent2].

and

(b) [antecedent1]'s [antecedent2], [antecedent1] hates [antecedent2].

4.1.1 No features differentiating the two competing antecedents

Both antecedents contain the same features. For simplicity of exposition, throughout this chapter we choose both antecedents to be masculine, singular, and third person in such circumstances, maintaining that the features themselves are irrelevant, as long as both competing antecedents are identical with respect to them. We know that the result would be just the same as if we had chosen, for instance, the feature combination feminine, plural, and third person\(^1\). For the sake of illustration, we present alternative structures that could be used to test the same conditions; these will be indicated in the template by primes.

Template: (1) John\(^2\)'s father\(^1\), (he\(^1\)) hates him\(^2\).
(2) John\(^1\)'s father\(^2\), (he\(^1\)) hates him\(^2\).
(1') Mary\(^2\)'s mother\(^1\), (she\(^1\)) hates her\(^2\).
(2') Mary\(^1\)'s mother\(^2\), (she\(^1\)) hates her\(^2\).

(1) shows co-reference with the topic antecedent, and is equivalent to Cole’s Type I.

(2) shows co-reference with the non-topic antecedent, and is equivalent to Cole’s Type K.

\(^1\) We did, of course, test for other appropriate combinations, such as matching feminine singular and plural antecedents, with no difference in the outcome of the grammatical judgements given. These are omitted here due to limitations of space.
We see above that all six varieties permit co-reference of pro with the topic 'father', and that none of them permits co-reference with the non-topic 'John'. Thus, no differences among the languages can be observed so far. We note here that all the (1) sentences are instances of same reference, and can license pro. This will hold consistently throughout the presentation of the data. The varying component will be the versions such as the (2) sentences, which contain switch reference.

4.1.2 Gender differentiating the two competing antecedents

Whenever the two competing antecedents do differ in respect to at least one feature, the template will list a few other possible combinations of these feature differences for illustration. However, we will only present data corresponding to the first pair of sentences in the template, repeating our stance that it is the feature differences, and not the actual selection of the features themselves, that account for the patterns.

---

2 We use in our Catalan examples the verb menyspreuar ('to despise'), because vowel-initial odiar ('to hate') can cause elision of the preceding clitic object, making it indistinguishable by gender in selected circumstances.
(3) shows co-reference with the topic antecedent, and is equivalent to the slot indicated by (3) in the table in Section 3.7.6.

(4) shows co-reference with the non-topic antecedent, and is equivalent to Cole’s Type J.

The only difference between (1) and (2), and (3) and (4), respectively, is that in the latter pair, the clitic is able to distinguish between the two antecedents present. This is Cole’s method of formalising competition of antecedents, the second condition in Ariel’s Accessibility Theory. In (1) and (2), both the topic and its possessor are potential candidates to serve as the antecedent to the subject. In (3) and (4), only one of these two is a logical option, thus reducing competition.

Catalan: (3ct) La mare₁ d’en Joan₂, (pro₁) el₂ menyspreua.  (4ct) La mare₂ d’en Joan₁, *pro₁/ell₁ la₂ menyspreua.

Castilian: (3cs) La madre₁ de Juan₂, (pro₁) le₂ odia.  (4cs) La madre₂ de Juan₁, *pro₁/él₁ la₂ odia.


Italian: (3it) La madre₁ di Gianni₂, (pro₁) lo₂ odia.  (4it) La madre₂ di Gianni₁, *pro₁/lui₁ la₂ odia.


P. R. Spanish: (3pr) La madre₁ de Juan₂, (pro₁) lo₂ odia.  (4pr) La madre₂ de Juan₁, *pro₁/el₁ la₂ odia.

The pattern arising from this set of data shows no differences whatsoever from that of Section 4.1.1. All the (3) sentences permit co-reference between pro and the topic, and all the (4) sentences require an overt subject pronoun for co-reference
between the subject and the non-topic antecedent. We are still unable at this point to
draw any distinctions among our languages tested.

4.1.3 **Number differentiating the two competing antecedents**

Template:  
(5) John’s neighbours, (they) hate him.  
(6) John’s neighbours, (he) hates them.  
(5') The sisters’ niece, (she) hates them.  
(6') The sisters’ niece, (they) hate her.

(5) is a modification of (1) but with differences in the number feature only.  
(6) is a modification of (2) but with differences in the number feature only.

Catalan:  
(5ct) Els veïns d’en Joan, (pro1) els menyspreuen.  
(6ct) Els veïns d’en Joan, *pro1/ell1 els menyspreuen.

Castilian:  
(5cs) Los vecinos de Juan, (pro1) le odian.  
(6cs) Los vecinos de Juan, *pro1/él1 los odia.

Eur. Port.:  
(5ep) Os vizinhos do João, (pro1) odeiam-no.  
(6ep) Os vizinhos do João, *pro1/el1 ele1 odeia-os.

Italian:  
(5it) I vicini di Gianni, (pro1) lo odiano.  
(6it) I vicini di Gianni, *pro1/lui1 li2 odia.

Braz. Port.:  
(5bp) Os vizinhos do João, (pro1) odeiam ele.  
(6bp) Os vizinhos do João, *pro1/el1 ele1 odeia eles.

P. R. Spanish:  
(5pr) Los vecinos de Juan, (pro1) lo odian.  
(6pr) Los vecinos de Juan, *pro1/el1 los odia.

Still no differences are observable among the data in the six varieties.
4.1.4 Person\textsuperscript{3} differentiating the two competing antecedents

Template: 
(7) My\textsubscript{2} father\textsubscript{1}, (he\textsubscript{1}) hated me\textsubscript{2}.
(8) My\textsubscript{1} father\textsubscript{2}, (I\textsubscript{1}) hated him\textsubscript{2}.
(7') Your\textsubscript{2} father\textsubscript{1}, (he\textsubscript{1}) hated you\textsubscript{2}.
(8') Your\textsubscript{1} father\textsubscript{2}, (you\textsubscript{1}) hated him\textsubscript{2}.

(7) is a modification of (1) but with differences in the person feature only.
(8) is a modification of (2) but with differences in the person feature only.

As most Romance varieties show reduced morphological distinctions in person within certain verbal paradigms, such as the imperfect and the conditional, we use these forms whenever they show such syncretism. This approach makes our methodology more rigorous because it causes the clitic to be the only linguistic unit able to distinguish between the two antecedents, since differences in verbal desinences are neutralised.

Catalan: 
(7ct) El meu\textsubscript{2} pare\textsubscript{1}, (pro\textsubscript{1}) em\textsubscript{2} menyspreuava.
(8ct) El meu\textsubscript{1} pare\textsubscript{2}, (pro\textsubscript{1}) el\textsubscript{2} menyspreuava.

Castilian: 
(7cs) Mi\textsubscript{2} padre\textsubscript{1}, (pro\textsubscript{1}) me\textsubscript{2} odiaba.
(8cs) Mi\textsubscript{1} padre\textsubscript{2}, (pro\textsubscript{1}) le\textsubscript{2} odiaba.

Eur. Port.: 
(7ep) O meu\textsubscript{2} pai\textsubscript{1}, (pro\textsubscript{1}) odiava-me\textsubscript{2}.
(8ep) O meu\textsubscript{1} pai\textsubscript{2}, (pro\textsubscript{1}) odiava-02.

Italian: 
(7it) Mio\textsubscript{2} padre\textsubscript{1}, (pro\textsubscript{1}) mi\textsubscript{2} odiava.
(8it) Mio\textsubscript{1} padre\textsubscript{2}, (pro\textsubscript{1}) lo\textsubscript{2} odiavo.

Braz. Port.: 
(7bp) O meu\textsubscript{2} pai\textsubscript{1}, (pro\textsubscript{1}) odiava-me\textsubscript{2}.
(8bp) O meu\textsubscript{1} pai\textsubscript{2}, (pro\textsubscript{1}) odiava ele\textsubscript{2}.

P. R. Spanish: 
(7pr) Mi\textsubscript{2} padre\textsubscript{1}, (pro\textsubscript{1}) me\textsubscript{2} odiaba.
(8pr) Mi\textsubscript{1} padre\textsubscript{2}, (pro\textsubscript{1})/yo\textsubscript{1} lo\textsubscript{2} odiaba.

\textsuperscript{3} Following Benveniste (1966), among others, we recognise three distinctions with regard to person: first (speaker), second (addressee), and third (non-participant in the discourse). We do not adopt the system of six different persons (first singular, second singular, third singular, first plural, second plural, third plural), instead choosing to divide up these six distinctions along two separate parameters, person as defined above and number (singular versus plural).
We now see our first instance of separation among the languages, albeit only one at the moment. Puerto Rican Spanish does not permit co-reference between a null subject and a non-topic antecedent when this antecedent differs from its topical counterpart by person specification only, as seen in (8pr). The other five varieties do permit such co-referencing. We hypothesize that this does not have anything to do directly with the ambiguous verbal morphology because we see identical patterns in Italian, which still distinguishes person in the imperfect, and in four of the varieties which do not. All six languages still mirror each other in instances of non-switch reference, as revealed by the (7) sentences.

This also brings initial awareness of the varying effects of the different features. In otherwise parallel circumstances, the person feature produces results not displayed by the gender or number features. Under our definition of person, it is not possible to compare two non-discourse participants with each other. At least one of the two must be a first person or a second person entity. This suggests that perhaps deictic factors are at least partially relevant. We expand further on this issue later.

4.1.5 Multiple features differentiating the two competing antecedents

It is possible for the two competing antecedents to differ with respect to more than one feature at a time. Four such combinations are possible. They may differ in gender and number only ((9) and (10)); in person and number only ((11) and (12)); in gender and person only ((13) and (14)); and in all three at once ((15) and (16)). For these types, (13) through (16) will seem contrived, since the more natural method of expressing possession would be to use a possessive instead of de plus a pronoun with a floating quantifier. However, these specially constructed possessive pronouns would appear very natural when used in the types involving two sentences (presented in Section 4.3 below), so we use them here to maintain consistency in the data presentation, without any hindrance to our argument.

Template:  
(9) Mary2’s neighbours1, (they1) hate her2.  
(10) Mary1’s neighbours2, (she1) hates them2  
(11) My2 neighbours1, (they1) hate me2.  
(12) My1 neighbours2, (I1) hate them2  
(13) Our2 (fem.) neighbours1, (they1) hate us2.
(14) Our1 (fem.) neighbours2, (we1) hate them2.
(15) Our2 (fem.) neighbour1, (he1) hates us2.
(16) Our1 (fem.) neighbour2, (we1) hate him2.

Catalan:

(9ct) Els veins1 de la Maria2, (pro1) la2 menyspreuen.
(10ct) Els veins2 de la Maria1, *pro1 / ella1 els2 menyspreua.
(11ct) Els meus2 veins1, (pro1) em2 menyspreuen.
(12ct) Els meus1 veins2, (pro1) els2 menyspreu.
(13ct) Els veins3 de totes nosaltres2, (pro1) ens2 menyspreuen.
(14ct) Els veins2 de totes nosaltres1, (pro1) els2 menyspreuem.
(15ct) El vein de totes nosaltres2, (pro1) ens menyspreu.
(16ct) El vein2 de totes nosaltres1, (pro1) el2 menyspreuem.

Castilian:

(9cs) Los vecinos1 de Maria2, (pro1) la2 odian.
(10cs) Los vecinos2 de Maria1, *pro1 / ella1 los2 odia.
(11cs) Mis2 vecinos1, (pro1) me2 odian.
(12cs) Mis1 vecinos2, (pro1) los2 odio.
(13cs) Los vecinos1 de todas nosotras2, (pro1) nos2 odian.
(14cs) Los vecinos2 de todas nosotras1, (pro1) los2 odiamos.
(15cs) El vecino1 de todas nosotras2, (pro1) nos2 odia.
(16cs) El vecino2 de todas nosotras1, (pro1) le2 odiamos.

Eur. Port.

(9ep) Os vizinhos1 da Maria2, (pro1) odeiam-na2.
(10ep) Os vizinhos2 da Maria1, *pro1 / ela1 os2 odia.
(11ep) Os meus2 vizinhos1, (pro1) odeiam-me2.
(12ep) Os meus1 vizinhos2, (pro1) odeio-os2.
(13ep) Os vizinhos1 de todas nós2, (pro1) odiam-nos2.
(14ep) Os vizinhos2 de todas nós1, (pro1) odiam-nos2.
(15ep) O vizinho1 de todas nós2, (pro1) odeia-nos2.
(16ep) O vizinho2 de todas nós1, (pro1) odiam-no2.

Italian:

(9it) I vicini1 di Maria2, (pro1) la2 odiano.
(10it) I vicini2 di Maria1, *pro1 / lei1 li2 odia.
(11it) I miei2 vicini1, (pro1) mi2 odiano.
(12it) I miei1 vicini2, (pro1) li2 odio.
(13it) I vicini1 di tutte noi2, (pro1) ci2 odiano.
(14it) I vicini2 di tutte noi1, (pro1) li2 odiamo.
(15it) Il vicino1 di tutte noi2, (pro1) ci2 odia.
(16it) Il vicino2 di tutte noi1, (pro1) lo2 odiamo.

Braz. Port.:  

(9bp) Os vizinhos1 da Maria2, (pro1) odeiam ela2.
(10bp) Os vizinhos2 da Maria1, *pro1 / ela1 odeia eles2.
(11bp) Os meus2 vizinhos1, (pro1) odeiam-me2.
(12bp) Os meus1 vizinhos2, (pro1) odeio eles2.
(13bp) Os vizinhos1 de todas nós2, (pro1) odeiam-nos2.
(14bp) *Os vizinhos_2 de todas nós_1, (pro_1) odiamos eles_2.
(15bp) *O vizinho_1 de todas nós_2, (pro_1) odeia-nos_2.
(16ep) *O vizinho_2 de todas nós_1, (pro_1) odiamos eles_2.

P. R. Spanish: (9pr) *Los vecinos_1 de María_2, (pro_1) la_2 odian.
(10pr) *Los vecinos_2 de María_1, *pro_1 / ella_1 los_2 odia.
(11pr) *Mis_2 vecinos_1, (pro_1) me_2 odian.
(12pr) *Mis_1 vecinos_2, *pro_1 / yo_1 los_2 odio.
(13pr) *Los vecinos_1 de todas nosotras_2, (pro_1) nos_2 odian.
(14pr) *Los vecinos_2 de todas nosotras_1, *pro_1 / nosotras_1 los_2 odiamos.
(15pr) *El vecino_1 de todas nosotras_2, (pro_1) nos_2 odia.
(16pr) *El vecino_2 de todas nosotras_1, *pro_1 / nosotras_1 lo_2 odiamos.

Note that in all six varieties, the (10) sentences, containing differences in gender and number between the potential antecedents, did not permit the coreferencing of *pro with the non-topic antecedent. We recall that in Sections 4.1.2 and 4.1.3, gender and number, respectively, were each shown to be unable on their own to license this as well. We observed in Section 4.1.4 that the person feature difference was enough to permit coreferencing between *pro and the non-topic antecedent in all languages except Puerto Rican Spanish. For these five varieties, (12), (14), and (16), are grammatical. What all of these three sentences have in common in each language is that they are instances of switch reference where the topic and non-topic antecedents differ with respect to at least the person feature. In Puerto Rican Spanish, (12pr), (14pr), and (16pr) do not allow *pro, and we saw in Section 4.1.4 that in this Romance variety, the person feature is not able to license *pro in this syntactic environment.

So as not to disrupt the flow of presentation, we no longer look at simultaneous combinations of more than one feature in any of the other types, claiming here that such combinations are able to license coreference in a certain language as long as at least one of the component features of that combination, on its own, can do so under the same syntactic conditions\(^4\). To illustrate, Catalan, for this type, is able to license coreference between *pro and a non-topic antecedent that differs from the topical antecedent with respect to both person and number (as well as to all three of person, number, and gender), because it can do so on the basis of

\(^4\) We have indeed tested for all such combinations for the sake of completeness, but we have chosen not to discuss all of them here, due to limitations of space and continuity of presentation.
differences in the person feature alone. We leave a more rigorous defence of our assumption for the analysis in later chapters.

4.2 **Topic followed by Main Clause and Subordinate Clause**

These sentences follow the format:

(c) [antecedent2]'s [antecedent1], we know that [antecedent1] hates [antecedent2].

and

(d) [antecedent1]'s [antecedent2], we know that [antecedent1] hates [antecedent2].

The principal difference between (c) and (d), and (a) and (b) from Section 4.1, respectively, is the inclusion of an intervening clause between the null subject pronoun and the antecedent candidates. This is Cole's attempt at formalising distance, the first component of Accessibility Theory. The antecedents in (c) and (d) are more remote from *pro* than their counterparts in (a) and (b), making them less accessible, according to Ariel.

### 4.2.1 No features differentiating the two competing antecedents

**Template:**

(17) John2's father1, we know that (he1) hates him2.
(18) John1's father2, we know that (he1) hates him2.
(17') Mary2's mother1, we know that (she1) hates her2.
(18') Mary1's mother2, we know that (she1) hates her2.

(17) shows co-reference with the topic antecedent, and corresponds to Cole's Type F.

(18) shows co-reference with the non-topic antecedent, and corresponds to Cole's Type H.

**Catalan:**

(17ct) _El pare1 d'en Joan2, sabem que (pro1) el2 menyspreua._
(18ct) _El pare2 d'en Joan1, sabem que *pro1 ell1 el2 menyspreua._

138
We observe no differences between (17) and (18), and their counterparts without the intervening main clause (1) and (2), respectively, in any part of the data. When there are no differentiating factors between the topic and the possessor, coreference between pro and the topic is still permitted, and an overt pronoun is still needed between the subject of the embedded clause and the non-topic antecedent.

We also note that (17) and (18) are consistent throughout the six varieties, so we are unable to draw any further distinctions among them.

4.2.2 Gender differentiating the two competing antecedents

Template: (19) John$_2$’s mother$_1$, we know that (she$_1$) hates him$_2$.
(20) John$_1$’s mother$_2$, we know that (he$_1$) hates her$_2$.
(19’$) Mary$_2$’s father$_1$, we know that (he$_1$) hates her$_2$.
(20’$) Mary$_1$’s father$_2$, we know that (she$_1$) hates him$_2$.

(19) shows co-reference with the topic antecedent, and corresponds to the slot indicated by (19) in the table in Section 3.7.6.

(20) shows co-reference with the non-topic antecedent, and corresponds to Cole’s Type G.

Catalan: (19ct) La mare$_1$ d’en Joan$_2$, sabem que (pro$_1$) el$_2$ menyspreua.
(20ct) La mare$_2$ d’en Joan$_1$, sabem que *pro$_1$/ell$_1$ la$_2$ menyspreua.
Castilian:  (19cs) La madre1 de Juan2, sabemos que (pro1) le2 odia.
(20cs) La madre2 de Juan, sabemos que *pro1/ él1 la2 odia.

Eur. Port.:  (19ep) A mãe1 do João2, sabemos que (pro1) o2 odeia.
(20ep) A mãe2 do João1, sabemos que *pro1/ ele1 ela2 odeia.

Italian:  (19it) La madre1 di Gianni2, sappiamo che (pro1) lo2 odia.
(20it) La madre2 di Gianni1, sappiamo che *pro1/ lui1 la2 odeia.

Braz. Port.:  (19bp) A mãe1 do João2, sabemos que (pro1) odeia ele2.
(20bp) A mãe2 do João1, sabemos que *pro1/ ele1 odeia ela2.

P. R. Spanish:  (19pr) La madre1 de Juan2, sabemos que (pro1) lo2 odia.
(20pr) La madre2 de Juan1, sabemos que *pro1/el1 la2 odia.

As was the case in the immediately preceding subsection, no differences are discernible between (19) and (20), and their shorter counterparts (3) and (4), respectively, which lack the intervening main clause. When gender is the only factor separating the two antecedents, co-reference between pro and the topic is tolerated, while an overt pronoun is required for the same between the subject of the embedded clause and the non-topic antecedent.

All language-specific versions of (19) and (20) are consistent with each other throughout the data, and hence do not help in distinguishing among them. Up to this point, only Puerto Rican Spanish differs from the other five, as we saw in Section 4.1.4 when looking at differences in person feature.

4.2.3 **Number differentiating the two competing antecedents**

Template:  (21) John1’s neighbours1, we know that (they1) hate him2.
(22) John1’s neighbours2, we know that (he1) hates them2.
(21’) The sisters2’ niece1, we know that (she1) hates them2.
(22’) The sisters1’ niece2, we know that (they1) hate her2.

(21) is a modification of (17) but with differences in the number feature only.
(22) is a modification of (18) but with differences in the number feature only.

Catalan:  (21ct) Els veins1 d’en Joan2, sabem que (pro1) el2 menspreuen.
(22ct) Els veins2 d’en Joan1, sabem que (pro1) els2 menspreuen.
We now witness a second group of data showing cross-linguistic differences. All six varieties still permit co-reference between pro and the topic antecedent. However, Catalan, Castilian, and European Portuguese now permit this as well with the non-topic antecedent, as seen in (22ct), (22cs), and (22ep), respectively. This is a departure from (19) and (20), where differences in gender only, all other things being equal, did not license pro. (21) and (22) show separation from their main clause-lacking counterparts in (5) and (6) with respect to Catalan, Castilian, and European Portuguese.

It is now possible to make an initial rough partition among the languages. We have Catalan, Castilian, and European Portuguese as currently the languages most accommodating to null subject licensing, followed by Italian and Brazilian Portuguese, which are less so, with Puerto Rican Spanish as the least pro-friendly of the six thus far. We note as well the indications above that the number and gender features do not operate in the same manner.

4.2.4 Person differentiating the two competing antecedents

Template:  
(23) My2 father1, we know that (he1) hated me2.  
(24) My1 father2, we know that (l1) hated him2.  
(23’) Your2 father1, we know that (he1) hated you2.  
(24’) Your1 father2, we know that (you1) hated him2.
(23) is a modification of (17) but with differences in the person feature only. (24) is a modification of (18) but with differences in the person feature only.

Catalan:  (23ct) * El meu₂ pare₁, sabem que (pro₁) em₂ menyspreuava.
          (24ct) * El meu₁ pare₂, sabem que (pro₁) el₂ menyspreuava.

Castilian:  (23cs) Mi₂ padre₁, sabemos que (pro₁) me₂ odiaba.
            (24cs) Mi₁ padre₂, sabemos que (pro₁) le₂ odiaba.

Eur. Port.:  (23ep) O meu₂ pai₁, sabemos que (pro₁) me₂ odiava.
             (24ep) O meu₁ pai₂, sabemos que (pro₁) o₂ odiava.

Italian:  (23it) Mio₂ padre₁, sappiamo che (pro₁) mi₂ odiava.
          (24it) Mio₁ padre₂, sappiamo che (pro₁) lo₂ odiavo.

Braz. Port.:  (23bp) O meu₂ pai₁, sabemos que (pro₁) me₂ odiava.
              (24bp) O meu₁ pai₂, sabemos que (pro₁) odiava ele₂.

P. R. Spanish:  (23pr) Mi₂ padre₁, sabemos que (pro₁) me₂ odiaba.
                (24pr) Mi₁ padre₂, sabemos que *pro₁/yo₁ lo₂ odiaba.

As we saw in Section 4.1.4, while all varieties permit pro to take the topic as the antecedent, only Puerto Rican Spanish does not permit co-referencing between a null subject and the possessor when the only feature difference between the two antecedents is person. So the intervening main clause causes no differences in the data between (23) and (24), and their corresponding shorter versions (7) and (8), respectively.

Based on our observations in Section 4.2, we can formulate a couple of interim hierarchies. For the six languages, they appear to range, in order of “decreasing pro-drop”, as follows: Catalan/Castilian/European Portuguese >> Italian/Brazilian Portuguese >> Puerto Rican Spanish. As for features, we see in Section 4.2.4 that differences in person feature seem to be the best able to tolerate pro, followed by number as viewed in Section 4.2.3, and finally, gender as the weakest licenser.
4.3 Two sentences

These sentences follow the format:

(e) [antecedent1] is going to meet [antecedent2] today. [antecedent1] hates [antecedent2].

and

(f) [antecedent1] is going to meet [antecedent2] today. [antecedent2] hates [antecedent1].

These are slight deviations from Cole’s corresponding sentences. His types follow the pattern:

(α) [antecedent1]’s [antecedent2] is terrible. [antecedent1/2] hates [antecedent2/1].

(Cole 2000: 139-140)

We choose to remodel (α) as (e) and (f) because potential complications arise when an adjective is present. Since the adjective “terrible” agrees in number with the topic of the first sentence, and possibly in gender as well, depending on the specific word chosen as the appropriate translation in each language, its presence would confound our investigation in two respects. First, the presence of an adjective would mean that the features of the subject are manifested in two locations, on the verb (in person and number, through Spec-head agreement) and on the adjective (with regard to gender and number features), so this could complicate our understanding of the effects of features in licensing the null subject (cf. Cole (2000: 139; 146)). (e) and (f) show subject features only on the verb with which it agrees. In this respect, (e) and (f) mirror the other sentences in Sections 4.1 and 4.2, whereas (α) deviates.

Second, no adjectives are present in the types reviewed up to this point, and therefore, to depart from the standard control sentences by using an inflecting adjective here in Section 4.3, when they were absent in Sections 4.1 and 4.2, would be methodologically inconsistent. Furthermore, we observe no tangible benefit gained
by our examination in having this adjective present. Therefore, we feel justified in using our versions of the sentences as stipulated in (e) and (f).

The two-sentence sequences (e) and (f) are used in comparison to (a) and (b), respectively, in Section 4.1, to highlight the influence of left-dislocation topicality, which is the third condition in Ariel’s Accessibility Theory. We may consider (a) and (b) to be roughly the topicalised equivalents of (e) and (f).

4.3.1 No features differentiating the two competing antecedents

Template:  
(25) John₁ is going to meet Mike₂ today. (He₁) hates him₂.  
(26) John₁ is going to meet Mike₂ today. (He₂) hates him₁.  
(25') Mary₁ is going to meet Laura₂ today. (She₁) hates her₂.  
(26') Mary₁ is going to meet Laura₂ today. (She₂) hates her₁.

(25) shows co-reference with the topic antecedent, and is the counterpart to Cole’s Type C. We note that it is an instance of same-reference.

(26) shows co-reference with the non-topic antecedent, and is the counterpart to Cole’s Type E. This is an example of switch reference.

Catalan:  
(25ct) En Joan₁ es reuneix amb en Miquel₂ avui. (pro₁) El₂ menyspreua.  
(26ct) En Joan₁ es reuneix amb en Miquel₂ avui. (pro₂) El₁ menyspreua.

Castilian:  
(25cs) Juan₁ va a quedar con Miguel₂ hoy. (pro₁) Le₂ odia.  
(26cs) Juan₁ va a quedar con Miguel₂ hoy. *pro₂ / El₂ le₁ odia.

Eur. Port.:  
(25ep) O Joao₁ encontra-se com o Miguel₂ hoje. (pro₁) Odeia-o₂.  
(26ep) O Joao₁ encontra-se com o Miguel₂ hoje. *pro₂ / Ele₂ oedia-o₁.

Italian:  
(25it) Gianni₁ incontra Michele₂ oggi. (pro₁) Lo₂ odia.  
(26it) Gianni₁ incontra Michele₂ oggi. *pro₁/ Lui₂ lo₁ odia.

Braz. Port.:  
(25bp) O Joao₁ se encontra com o Miguel₂ hoje. (pro₁) Odea ele₂.  
(26bp) O Joao₁ se encontra com o Miguel₂ hoje. *pro₂ / Ele₂ odeia ele₁.

P. R. Spanish:  
(25pr) Juan₁ va a encontrarse con Miguel₂ hoy. (pro₁) Lo₂ odia.  
(26pr) Juan₁ va a encontrarse con Miguel₂ hoy. *pro₂ / El₂ lo₁ odia.
Unsurprisingly, all six varieties permit pro in instances of same reference, as sentences (25) reveal. However, what is noteworthy is that only Catalan still licenses co-reference between pro and a non-topic antecedent, as seen in (26ct). The corresponding sentences in the other five languages require an overt subject pronoun. We now have a means by which to separate Catalan from Castilian and European Portuguese, and hypothesise at this point that Catalan is the most pro-drop of the varieties under investigation. It remains to be seen whether this conclusion holds up in the data that follow.

4.3.2 Gender differentiating the two competing antecedents

Template:  
(27) John₁ is going to meet Mary₂ today. (He₁) hates her₂.  
(28) John₁ is going to meet Mary₂ today. (She₂) hates him₁.  
(27') Laura₁ is going to meet Mike₂ today. (She₁) hates him₂.  
(28') Laura₁ is going to meet Mike₂ today. (He₂) hates her₁.

(27) shows co-reference with the topic antecedent, and corresponds to the slot indicated by (27) in the table in Section 3.7.6. (27) is an instance of same reference.

(28) shows co-reference with the non-topic antecedent, and is the counterpart to Cole’s Type D. This is an example of switch reference.

Catalan:  
(27ct) En Joan₁ es reuneix amb la Maria₂ avui. (pro₁) La₂ menyspreua.  
(28ct) En Joan₁ es reuneix amb la Maria₂ avui. (pro₂) El₁ menyspreua.

Castilian:  
(27cs) Juan₁ va a quedar con Maria₂ hoy. (pro₁) La₂ odioa.  
(28cs) Juan₁ va a quedar con Maria₂ hoy. (pro₂) Le₁ odioa.

Eur. Port.:  
(27ep) O João₁ encontra-se com a Maria₂ hoje. (pro₁) Odeia-a₂.  
(28ep) O João₁ encontra-se com a Maria₂ hoje. *pro₂ / Ela₂ odeia-o₁.

Italian:  
(27it) Gianni₁ incontra Maria₂ oggi. (pro₁) La₂ odioa.  
(28it) Gianni₁ incontra Maria₂ oggi. *pro₂ / Lei₂ lo₁ odioa.

Braz. Port.:  
(27bp) O João₁ se encontra com a Maria₂ hoje. (pro₁) Odeia ela₂.  
(28bp) O João₁ se encontra com a Maria₂ hoje. *pro₂ / Ela₂ odeia ele₁.

P. R. Spanish:  
(27pr) Juan₁ va a encontrarse con Maria₂ hoy. (pro₁) La₂ odioa.  
(28pr) Juan₁ va a encontrarse con Maria₂ hoy. *pro₂ / Ella₂ lo₁ odioa.
When feature differences with regard to gender are manifested, both Catalan and Castilian support the switch reference, permitting co-reference between pro and a non-topic antecedent as seen in (28ct) and (28cs) respectively. The other four varieties require an overt pronoun for co-reference between the subject of the second sentence and the non-topic antecedent. This does not contradict our language hierarchy with Catalan and Castilian at the top. All instances of same reference as exemplified in the (27) sentences allow pro.

4.3.3 Number differentiating the two competing antecedents

Template: (29) John₁ is going to meet the neighbours₂ today. (He₁) hates them₂.
(30) John₁ is going to meet the neighbours₂ today. (They₂) hate him₁.
(29') Mary₁ is going to meet the sisters₂ today. (She₁) hates them₂.
(30') Mary₁ is going to meet the sisters₂ today. (They₂) hate her₁.

(29) is a modification of (25) but with differences in the number feature only.
(30) is a modification of (26) but with differences in the number feature only.

Catalan: (29ct) En Joan₁ es reuneix amb els veïns₂ avui. (pro₁) Els₂ menyspreua.
(30ct) En Joan₁ es reuneix amb els veïns₂ avui. (pro₂) El₁ menyspreuen.

Castilian: (29cs) Juan₁ va a quedar con los vecinos₂ hoy. (pro₁) Los₂ odia.
(30cs) Juan₁ va a quedar con los vecinos₂ hoy. (pro₂) Le₁ odian.

Eur. Port.: (29ep) O João₁ encontra-se com os vizinhos₂ hoje. (pro₁) Odeia-os₂.
(30ep) O João₁ encontra-se com os vizinhos₂ hoje. (pro₂) Odeiam-no₁.

Italian: (29it) Gianni₁ incontra i vicini₂ oggi. (pro₁) Li₂ odia.
(30it) Gianni₁ incontra i vicini₂ oggi. (pro₂) Lo₁ odiano.

Braz. Port.: (29bp) O João₁ se encontra com os vizinhos₂ hoje. (pro₁) Odeia eles₂.
(30bp) O João₁ se encontra com os vizinhos₂ hoje. *pro₂ / Elles₂ odeiam ele₁.

P. R. Spanish: (29pr) Juan₁ va a encontrarse con los vecinos₂ hoy. (pro₁) Los₂ odia.
(30pr) Juan₁ va a encontrarse con los vecinos₂ hoy. *pro₂ / Ellos₂ lo₁ odian.
Greater differences among the six varieties are manifested in this group of data. Catalan, Castilian, European Portuguese, and Italian permit co-referencing between *pro* and the non-topic antecedent when the two competing antecedents differ in number only. Brazilian Portuguese (30bp) and Puerto Rican Spanish (30pr) require an overt pronoun to enable co-referencing between the subject of the second sentence and the non-topic antecedent in the first.

These data are able to separate Italian from Brazilian Portuguese, which had been indistinguishable up to this point in our investigation. Thus, the status of our language hierarchy is as follows:

Catalan >> Castilian >> European Portuguese >> Italian >> Brazilian Portuguese >> Puerto Rican Spanish

The differences in the data between Section 4.3.2 and Section 4.3.3 suggest that the appearance of the number feature is more liberal than that of gender in the licensing of the null subject, all other things being equal. Finally, the instances of same reference in the six varieties remain undiscriminating, as seen in the (29) sentences allowing *pro*. This further supports our justification for not using Cole's Type A in the present examination.

4.3.4 Person differentiating the two competing antecedents

Template:  
(31) (I₁) am going to meet my father₂ today. (I₁) used to hate him₂.  
(32) (I₁) am going to meet my father₂ today. (He₂) used to hate me₁.  
(31') (We₁) are going to meet the neighbours₂ today. (We₁) used to hate them₂  
(32') (We₁) are going to meet the neighbours₂ today. (They₂) used to hate us₁.

(31) is a modification of (25) but with differences in the person feature only.  
(32) is a modification of (26) but with differences in the person feature only.
We see that the only divergence in the data is the inability in Puerto Rican Spanish to permit co-reference between pro in the second sentence and the non-topic antecedent mi padre in the first, as shown in (32pr). All other versions of (32) do permit such co-reference when number only differentiates between the two competing antecedents of the first sentence. In light of this data set, then, our current ranking of languages, in decreasing order of pro-drop, does not need to be amended, remaining as follows:

Catalan >> Castilian >> European Portuguese >> Italian >> Brazilian Portuguese >> Puerto Rican Spanish

We also take this opportunity to point out that switch reference was permitted in two languages (Catalan and Castilian) when gender was the only factor separating the competing antecedents, in four languages (the previous two, in addition to European Portuguese and Italian) when number was the only feature difference, and in five languages (all but Puerto Rican Spanish) when person was the sole feature. We can thus suggest a hierarchy of features, according to how well their presence as a
distinguishing feature permits the null subject. This ranking, in decreasing order of strength of pro-drop licensing, would be person >> number >> gender.

Everything looks to be in order. However, at this point we present data that may be an obstacle to the two hierarchies just mentioned. After examining these results, we return to our hierarchies and investigate whether they need to be amended.

4.4 Gender-inherent predicates

Cole does not examine Ariel's fourth condition in her Accessibility Theory, unity, with the same meticulousness as he does with the other three. In Cole's defence, it is admitted that this fourth component seems much more difficult to investigate objectively than distance, topicality, and competition. We attempt to examine unity here, in accordance with our interpretation of this term as presented in Section 3.7.4 to mean our knowledge of the real world. This will be done by resorting to the use of data on gender-inherent predicates.

Our fundamental claim is that the semantics of a clause will make unambiguous the intended antecedent of its subject. In particular, when the predicate of a clause stipulates that its logical subject should be feminine, we assert that if two competing antecedents are available, one masculine and the other feminine, then our grasp of reality, along with the natural, default interpretation of the clause in a non-imaginary context, will clearly bias one antecedent candidate over the other. In other words, if our predicate involves being pregnant, and we have no motivation to assume that this statement should be interpreted with respect to a parallel universe where men possess the ability to bear children, then the feminine antecedent should be dominantly favoured over the male antecedent.

We check the validity of this assumption in the following manner. First, we consider that gender may be indicated in a predicate through both morphological agreement and through the semantics of the clause, and that these two are not necessarily mutually independent. Therefore, there are four ways in which the preferred gender may be specified: [+semantic, +morphological], [+semantic, -morphological], [-semantic, +morphological], and [-semantic, -morphological]. It should be pointed out that this fourth combination, consisting of two negative settings,
is actually the lack of gender indication in the predicate. Thus, it serves as the basis upon which the other three combinations are compared; it is the default case.

We define semantic indication of gender as being a statement where we use our knowledge of reality to infer that the only possible subject of the statement is feminine, for example, someone giving birth. Morphological identification of gender consists of the feminine version of an adjective, regardless of whether a masculine version exists. This can be illustrated by the use of Italian feminine bella, ‘beautiful’, where a masculine version, bello, is grammatical and can be used to indicate the same descriptive quality. This can also be exemplified by Italian incinta, ‘pregnant’, even though the masculine equivalent, incinto, is not generally used\(^5\). The use of incinta would additionally satisfy the conditions on semantic identification, whereas bello/a would not.

For each of the four setting combinations, we look at both agentive and stative predicates, for a total of eight different items under examination. We hypothesised in Chapter 2 that the thematic role has no direct bearing on the ability to license pro. We use the exploration of this data either to confirm or to amend our position.

The table below outlines which predicates correspond to which of the eight combinations:

<table>
<thead>
<tr>
<th>FACTOR SETTINGS</th>
<th>GENDER-STATIVE</th>
<th>GENDER-AGENTIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>[+sem, +morph]</td>
<td>to be eight months pregnant (33)</td>
<td>to establish oneself as prima ballerina (37)</td>
</tr>
<tr>
<td>[+sem, -morph]</td>
<td>to give birth (34)</td>
<td>to buy oneself maternity clothes (38)</td>
</tr>
<tr>
<td>[-sem, +morph]</td>
<td>to be beautiful (35)</td>
<td>make oneself look beautiful (39)</td>
</tr>
<tr>
<td>[-sem, -morph]</td>
<td>to feel ill (36)</td>
<td>to behave well (40)</td>
</tr>
</tbody>
</table>

The English sentence templates for these eight categories are as follows:

---

\(^5\) There are a few exceptions to this, such as in the title of a 1973 Italian film starring Marcello Mastroianni, *Niente di grave, suo marito è incinto.*

150
(33) John is taking Mary to the hospital today. (She) is eight months pregnant.

(34) John is taking Mary to the hospital today. (She) is about to give birth.

(35) John is now dating Mary. (She) is very beautiful.

(36) John is taking Mary to the hospital today. (She) still feels ill.

(37) John is now dating Mary. (She) has recently established herself as the prima ballerina of this country.

(38) John is taking Mary to the shopping mall today. (She) is going to buy herself some new maternity clothes.

(39) John has been waiting for Mary for quite a while now. (She) is busy making herself beautiful for the ball.

(40) John can take Mary out to the supermarket. (She) behaves well in public.

Translated versions, in each of the six Romance varieties under consideration, were tested for whether *pro* was permitted at the beginning of the second sentence in each item, or if the overt feminine subject pronoun was required to render the sentence grammatical. The results are presented below; the actual sentences appear in Appendix C.

**LANGUAGES THAT PERMIT *pro* IN THE FOLLOWING SENTENCES**

<table>
<thead>
<tr>
<th>FACTOR SETTINGS</th>
<th>GENDER-STATIVE</th>
<th>GENDER-AGENTIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>[+sem, +morph]</td>
<td>Cast, Cat, Italian</td>
<td>Cast, Cat, Italian</td>
</tr>
<tr>
<td>[+sem, -morph]</td>
<td>Cast, Cat, Italian</td>
<td>Cast, Cat, Italian</td>
</tr>
<tr>
<td>[-sem, +morph]</td>
<td>Cast, Cat, Italian</td>
<td>Cast, Cat, Italian</td>
</tr>
<tr>
<td>[-sem, -morph]</td>
<td>Cat</td>
<td>Cat</td>
</tr>
</tbody>
</table>

We note that Catalan is the sole language that tolerates *pro* in all eight circumstances under investigation. In particular, it is the only variety able to license
pro in instances of switch reference, as seen in items (36) and (40), where there is no gender-inherent marking. This is consistent with our findings in Section 4.3.2, as well as with our language hierarchy where Catalan is established at its apex.

We see in the table that Castilian and Italian will license pro in switch reference in all six categories where there is some form of gender indication in the predicate, so they are clearly underneath Catalan on our language hierarchy. The potential pitfall, however, is that they should thus rank above the three varieties that do not tolerate pro in these conditions, European Portuguese, Brazilian Portuguese, and Puerto Rican Spanish. In particular, we draw the reader back to the hierarchy as presented at the end of Section 4.3, reproduced below:

Catalan >> Castilian >> European Portuguese >> Italian >> Brazilian Portuguese >> Puerto Rican Spanish

There is a possible inconsistency in the order of European Portuguese and Italian. Based on the data presented in Sections 4.1, 4.2, and 4.3, European Portuguese clearly ranks above Italian. However, what we see here in Section 4.4 shows that the opposite order, with Italian taking priority over European Portuguese, should apply. We are thus left with a quandary involving the ambiguous relative status of the languages intermediate in our ranking scheme. We examine possible resolutions below.

4.5 Exploratory generalisations

One potential solution is to declare Italian and European Portuguese on equal terms, citing the need for future research and additional data in order to separate the two. Thus the language hierarchy would be as follows, with European Portuguese and Italian occupying a joint-third position out of the six:

Catalan >> Castilian >> European Portuguese/Italian >> Brazilian Portuguese >> Puerto Rican Spanish
This is not a bad answer to the problem. However, we maintain that upon further exploration of the data presented in this chapter, it may be an unnecessary concession. A better resolution may exist that does not force such an amendment of our established rankings.

We draw attention to the fact that gender as indicated in Sections 4.1, 4.2, and 4.3 is different from the manner in which gender is specified in Section 4.4. In particular, the situation in the former identified gender on the basis of a clitic pronoun, whereas gender in the latter group of data was manifested by adjectival morphology. We propose, then, an interim separation of the gender feature into two separate entities, gender indicated by clitic and gender indicated by inflectional morphology. We recall from the summaries of Gilligan (1986) and Samek-Lodovici (1996) the acknowledgement of differences between clitics and synthetic inflectional desinences, although this idea was not greatly developed in those two works. Therefore, the suggested resolution presented here is not completely unfounded, although it still requires a formal justification in the following chapters.

If we amend our feature hierarchy so as to accommodate this revision of gender, we now have ranking systems involving four features: gender indicated by clitic, gender indicated by inflectional morphology, person, and number. Furthermore, depending on which version of the gender feature we use, we get two different language hierarchies. This is summarised below:

**Bundle 1** (with gender indicated by clitic):

Feature hierarchy: person >> number >> gender by clitic

Language hierarchy: Catalan >> Castilian >> European Portuguese >> Italian >> Brazilian Portuguese >> Puerto Rican Spanish

**Bundle 2** (with gender indicated by inflectional morphology):

Feature hierarchy: person >> number >> gender by inflectional morphology

Language hierarchy: Catalan >> Castilian >> Italian >> European Portuguese >> Brazilian Portuguese >> Puerto Rican Spanish

153
The first bundle reflects the ordering of languages when differences in gender are indicated by clitics, as reflected by the data in Section 4.1 through Section 4.3. The second bundle shows the language ranking when gender differences are readable from inflectional affixes, as seen in Section 4.4.

A question that arises, then, is whether it is efficient or reasonable to have two different bundled versions of these two hierarchies. One possible method of addressing this concern would be to merge the two bundles of hierarchies into one. This could be done in the following manner:

**Bundle 3** (with two different gender features, one for clitics, one for inflectional morphology):

Feature hierarchy: person >> number >> gender by clitic >> gender by inflectional morphology

Language hierarchy: Catalan >> Castilian >> European Portuguese >> Italian >> Brazilian Portuguese >> Puerto Rican Spanish

A necessary proviso would have to be appended to this bundle of hierarchies, stipulating that the higher (i.e., more leftward) we are on the feature hierarchy, the closer we observe the ordering of its companion hierarchy. For example, gender by clitic is higher than gender by inflectional morphology on the feature hierarchy, so the language hierarchy is obeyed better when we consider the data involving gender by clitic than when we consider the data with gender indicated by morphology. We can also note that the language hierarchy is no worse observed when we view the data involving person and number, features ranking higher on that hierarchy. This is admittedly a complicated amendment and hence an imperfect solution. We explore in subsequent chapters more preferable methods of resolving this ambivalence.

It is worth mentioning briefly here that different variants of the “same language” may actually be further apart than previously thought. Some may find it quite unexpected that Castilian is closer to European Portuguese than it is to Puerto Rican Spanish, its Caribbean counterpart, at least with regard to the null subject. Similarly, it may be surprising that European Portuguese shows greater affinity to Italian than it does to the corresponding Southern American version of the language. While we did not attempt a formal *a priori* justification for using what some may
consider “dialects of the same language”, we hope that the data presented in this chapter have established the validity of treating the distinct variations of Spanish and Portuguese as separate entities.

We have now provided an initial summary and framework of the information gathered from the collected data. This sets us up for a thorough examination of the possible analyses to explain these patterns, proper defences in support of them, and the broader conclusions reached in the present research. This is where we direct our attention in the next chapter.
Chapter 5: The validity of PPT approaches

We now attempt a thorough evaluation of the results presented in Chapter 4 and seek to provide an account of their ramifications on linguistic theory and on our understanding of the null subject phenomenon in the Romance languages. We begin by appraising here the success of those frameworks traditionally proffered to explain cross-linguistic variation in null subject usage.

5.1 Limitations of the Government and Binding framework

We present here our assessment of how well various frameworks address the null subject phenomenon by first examining the Government and Binding (henceforth GB) model within the Chomskyan Principles and Parameters Theory (henceforth PPT). It was argued in Section 1.1 that oversimplifications abound when it comes to explaining what constitutes a "pro-drop language" and, in particular, when subject pronouns in such languages have to be, or are normally, used. The situation is similar when we consider various definitions of the so-called "pro-drop parameter". Not only do these characterisations end up underspecifying how subject pronouns are employed in languages, but they are also largely inaccurate descriptions of what is really happening. This is a troubling prospect if we are to assume that these overgeneralised definitions of the "pro-drop parameter", which are flawed and thus not sound enough to support subsequent conclusions, will eventually serve as the foundation upon which we construct advanced, more detailed representations of subject pronoun usage.

5.1.1 Seeking out this alleged "pro-drop parameter"

5.1.1.1 Chomsky’s definitions

Chomsky (1981) was one of the first linguists to consider in depth the nature of a pro-drop parameter. We have already demonstrated in Section 1.2.1, and have
seen in the work of Gilligan (1987) as outlined in Section 3.2.1, that the bundling of certain syntactic properties, such as free inversion in simple sentences and long wh-movement of the subject, with the availability of covert subject pronouns is not completely correct. However, it is also worth examining how Chomsky actually defines the pro-drop parameter itself. He writes:

"[...] there is a single parameter of core grammar – the "pro-drop parameter" – that distinguishes Italian-type from French-type languages. [...] Let us assume – following Taraldsen (1978) – that the parameter involves the inflectional element INFL, or more precisely, the agreement element AGR (= PRO) that is the crucial component of INFL with respect to government and binding. The intuitive idea is that where there is overt agreement, the subject can be dropped, since the deletion is recoverable."

(Chomsky 1981: 241)

Chomsky thus correlates the positive setting of this pro-drop parameter with the nature of AGR and hence the rich verbal inflection available in a language. We now know, of course, that this is not the case. Our review of the quantitative survey offered by Gilligan (1987) in Section 3.2.2 and of the analysis of the situation in Icelandic given by Sigurðsson (1993) in Section 3.3 makes a strong argument against any direct linking between null subject usage and AGR. It is acknowledged, however, that Chomsky himself admits that this correlation is not an absolute principle but rather just a guiding inclination. He claims that:

"The correlation with overt inflection need not be exact. We expect at most a tendency in this direction. The idea is, then, that there is some abstract property of AGR, correlated more or less with overt morphology, that distinguishes pro-drop from non-pro-drop languages [...]"

(Chomsky 1981: 241)

Although this does not change the fact that Chomsky has still erred by asserting any existence whatsoever of a formal relation between overt verbal morphology and covert subject pronouns, we do grant that, from even the earliest stages of research on the null subject, linguists have recognised that it was not the case that subject pronouns could be absent only if verbal inflection was rich enough,
or that strong verbal desinences were sufficient to license subject *pro*. In other words, it has long been known, if not necessarily always clearly or explicitly stated, that there was more to the null subject phenomenon than merely the amount of variation present in a language’s verbal paradigm.

It should be noted that Chomsky does allude to the possibility of intra-linguistic variation in null subject usage, a consideration largely ignored by a number of later linguists who do not touch upon this at all in their descriptions of the *pro*-drop parameter. Chomsky, foreshadowing the research conducted by Gilligan (1987), Samek-Lodovici (1996), and Cole (2000) as detailed in Sections 3.2.3, 3.4, and 3.7, respectively, speculates that:

“...A language might have a mixed system, permitting subject drop in some constructions but not in others, a property that we might expect to find varying as inflection is or is not overt; Taraldsen [1978] gives examples from Irish; Hebrew is another case.”

(Chomsky 1981: 241)

Despite the unfortunate characterisation that language-internal variation in null subject usage is dependent upon the level of verbal inflection, Chomsky’s hypothesis that, within any one language, the types of construction have some sort of effect on the licensing of the null subject, in that particular linguistic environment, is a roughly accurate prediction of the results discovered by subsequent research. We now know that all six Romance varieties under investigation in the present work can be fairly well characterised by Chomsky’s use of the term “mixed system”.

5.1.1.2 Jaeggli and Safir’s (1989a) “morphological uniformity”

Jaeggli and Safir’s (1989a) definition of the null subject parameter focuses on verbal inflection as well, but their approach differs from Chomsky’s in that they place much less significance on the richness of verbal inflection, instead alluding to the type of verbal inflection as the key factor determining the licensing of null subjects. After

---

1 Raposo (1989: 278ff.) *inter alia* shows this quite convincingly in his work on the European Portuguese inflected infinitive.
Jaeggli and Safir (1989a: 27-29) argue that rich AGR and null subjects have nothing directly to do with each other, citing evidence from Spanish, German, Japanese, and McCloskey and Hale’s (1984) work on Irish, they postulate the following definition of the null subject parameter:

"Null subjects are permitted in all and only languages with morphologically uniform inflectional paradigms"  (Jaeggli and Safir 1989a: 29)

where morphological uniformity is defined as follows:

"An inflectional paradigm P in a language L is morphologically uniform [if and only if] P has either underived inflectional forms or only derived inflectional forms."  
(Jaeggli and Safir 1989a: 30)

Whereas Chomsky identifies the richness of verbal inflection as the crucial component in the licensing of null subjects, Jaeggli and Safir argue alternatively that the nature of a language’s verbal paradigm is a better indicator. Jaeggli and Safir’s proposal was an improvement over Chomsky’s; the former could account for the null subjects attested in languages such as Chinese and Japanese with (uniformly) underived inflectional forms, as well as languages with richer AGR, such as German, that do not allow null thematic subjects, something that could not be accommodated within Chomsky’s model.

However, Jaeggli and Safir’s rendition of the null subject parameter is not perfect, either. Afrikaans is, like Chinese (Huang 1984), a language with morphologically uniform inflectional paradigms as there is no overt inflectional marking (Biberauer 2003: 10), but subject pronouns are nevertheless required. If we are to consider that the first and third person singular forms of the Portuguese inflected infinitive are identical to the impersonal infinitive, then under Jaeggli and Safir’s definition, Portuguese should be classified as a language lacking morphologically uniform paradigms, and thus should not be able to license pro. We know that this is not the case (e.g., Raposo 1987: 86).

Additionally, the relevance of morphological uniformity to the null subject phenomenon is a suspect one. There is no a priori intuitive or theoretical reason why we should expect only those languages that are either extremely varied or completely
impoverished in their verbal paradigms to support covert subject pronouns. To claim that only the languages located on the extremes of a postulated "strength of AGR continuum" will license pro in the subject position does not really provide us any useful insight. Furthermore, counterexamples to this hypothesis are easily identified, among them Afrikaans on the weak end, and Icelandic on the strong side. Furthermore, this definition is still inconsistent with the situations in the standard and the Bavarian German varieties, both of which rest on the same place on this AGR continuum but differ greatly in null subject usage (Gilligan 1987: 170). A final inconsistency to be pointed out is that four of the six Romance varieties we are looking at in the present work have morphologically uniform rich inflectional paradigms and thus qualify under Jaeggli and Safir's definition of a null subject language, yet each of those four shows different pockets of resistance to the licensing of the null subject in certain syntactic environments, so it cannot be claimed that they observe the null subject parameter in identical manner to each other. Thus, there does not appear to be anything in Jaeggli and Safir's statement to accommodate the reality of intra-linguistic "mixed systems" as stipulated by Chomsky.

The formulations of the "null subject parameter" above rely heavily on the richness, presence, or absence of AGR; this is clearly a faulty approach in view of our current knowledge of the independence between the licensing of pro and verbal morphology. We have already mentioned that Chinese and Japanese, for instance, show no verbal desinences for number and person, yet tolerate the null subject, whereas Icelandic has a highly rich system of verbal agreement, but does not allow null subjects. Furthermore, even within only the context of languages with rich AGR, we see that richer 'rich AGR' languages are not necessarily friendlier to pro-drop than poorer 'rich AGR' languages. One may argue quite reasonably, for instance, that AGR is richer in Italian than it is in Castilian, based primarily on the fact that the first person singular in the Italian imperfect and conditional is unique (e.g., parlavo 'I was speaking' and parlava 'He was speaking'), whereas the same in Castilian is syncretic with the third person singular (e.g., Castilian hablaba 'I/He was speaking'). All other things are roughly equal: both languages contain six different representations in the present, preterit, and future tenses, both possess two grammatical forms in the imperfect subjunctive (e.g., Italian first and second person singular parlassi and third

---

2 This is the case if we accept the argument stated above that the inflectional paradigms in Brazilian and European Portuguese are not morphologically uniform because of the inflected infinitive.
person singular parlasse; Castilian first and third person singular hablara and second person singular hablaras), and it may be claimed as well that they each show one syncretism in the present subjunctive (Castilian hable and hablas; Italian parli and tu parli\(^3\)). In light of the above, one may fairly conclude that Italian has slightly richer AGR in its verbal paradigm than Castilian. Our data and hierarchy in Chapter 4 demonstrate, however, that Castilian is unequivocally more tolerant of pro than Italian. This result is the opposite of what we would expect under standard treatments of the licensing of pro.

5.1.1.3 Rizzi (1986)

Thus, a more accurate representation of null subjecthood would have to distance itself from all verbal inflectional considerations. One such approach is that offered by Rizzi (1986). Rizzi acknowledges that the traditional assumptions of pro being linked to strong verbal agreement provide an insufficient account of null elements, citing in support of this argument his own work on the existence of object pro in Italian but not English, despite both of these languages lacking object agreement on the verb (1986: 518). He proposes the following parameter as an alternative:

“pro is governed by \(X^0_Y\). […] The defining property of null subject languages can be looked at as a particular setting of [this] parameter. […] [I]n Italian, Spanish, etc., but not in English, French, etc., a governing Infl capable of assigning nominative Case is a member of \(X^0_Y\). […] We would then conclude that the class \(X^0_Y\) includes both Infl and V in Italian, whereas it is empty in English, a grammatical system that does not license any occurrence of pro. […] French is not a null-subject language, but it allows pro in object position […] pro can occur as a prepositional object in French; therefore, in this grammatical system \(X^0_Y = \{V, P\}\). In short, we may expect natural languages to vary from a maximally restrictive setting (no head is a possible licenser) to a maximally liberal setting (every head is a possible licenser).”

(Rizzi 1986: 519)

Two noteworthy features of this hypothesis deserve comment. The first is that this is a significant break from linking null subject tolerance to AGR; Rizzi is arguing

---

\(^3\) This is assuming that we adopt Cordin and Calabrese’s (1988: 540) postulation that overt tu is always required when the second person singular subject is intended (as explored in Section 1.3.2).
here instead that the principal factor controlling cross-linguistic variation in pro-drop is the number and type of elements in each language that may govern pro. V and Infl both license pro in Italian, so subject and object pro are permitted. In French, V and P, but not Infl, are acceptable governors; thus, object and prepositional pro, but not subject pro, is allowed. No element is able to govern pro in English, hence the absence of null arguments in that language. While commendable in its efforts in addressing the irrelevance of verbal morphology to the licitness of pro, this approach does not adequately resolve the problem of how to explain intra-linguistic variation. Rizzi’s proposal that Infl and V are proper governors of pro in Italian does not satisfactorily address the microvariation in our Italian data in Chapter 4, since there are clear instances when neither Infl nor V is able to govern subject pro (e.g., (4i): La madre di Gianni, *pro1 / lui1 la2 odia.).

The second observation has to do, in particular, with the final sentence of the Rizzi excerpt above. Rizzi anticipates that there is a wide choice of parameter settings available to languages that determine how pro-friendly they are, ranging from “maximally restrictive” on one end to “maximally liberal” on the other side, and presumably, numerous intermediary options between these two extremes. We question below in greater detail whether this variation is actually best constructed as a parameter. Furthermore, inherent in Rizzi’s description is the concept of a continuum, and this suggests that perhaps a fluid model incorporating hierarchical rankings and flexible orderings would be more successful in explaining the Romance null subject than any framework involving discrete and absolute parameters.

5.1.1.4 Contemporary descriptions of null subject usage

More recent definitions of a pro-drop parameter are not necessarily any more refined or developed than Chomsky’s, Jaeggli and Safir’s, or Rizzi’s proposals from the 1980s; some of them, in fact, may be even less accurate. Consider this sampling of definitions of the pro-drop parameter found in one English grammar and two introductory syntax textbooks, all from the 1990s:
“The parameter distinguishing Italian from English, or more generally the parameter which distinguishes languages in which the subject pronoun in finite clauses may be non-overt from those in which it cannot be non-overt, is called the pro-drop parameter. Italian is said to be a pro-drop language: it allows the subject pronoun to be dropped.”

(Haegeman and Guéron 1999: 597)

“There appears to be parametric variation between languages as to whether or not they allow finite verbs to have null subjects. The relevant parameter (termed the null subject parameter) would appear to be a binary one, with only two possible settings, viz. does/doesn’t allow finite verbs to have null subjects. There appears to be no language which allows the subjects of some finite verbs to be null, but not others – e.g., no language in which it is OK to say Drinks wine (meaning ‘He/she drinks wine’) but not OK to say Eats pasta (meaning ‘He/she eats pasta’). The range of grammatical variation found across languages appears to be strictly limited: there seem to be just two possibilities – languages either do or don’t systematically allow finite verbs to have null subjects (i.e., to have an understood subject which is not overtly expressed).”

(Radford 1997: 17)

“Italian differs from English in that the former, though not the latter, allows the subject of a finite clause to remain unexpressed. The parameter which distinguishes languages like English which do not allow a subject pronoun to be omitted and those like Italian which do is referred to as the pro-drop parameter.”

(Haegeman 1994: 19)

We can identify a serious technical concern within those definitions that must be addressed if we are to attempt a proper scrutiny of how the PPT framework handles the null subject phenomenon. The question that arises is whether pro-drop can be accurately characterised as a binary case with two mutually exclusive settings; in other words, whether we can truly say that either a language can drop the subject pronoun, or it cannot. Two of the three definitions above incorporate an implied either-or approach to this issue, and the characterisation offered by Radford, in fact, states this binary assumption clearly and unambiguously.

If we look at the data of any one particular language from Chapter 4, it becomes readily apparent that to label that language as being either a pro-drop language or not a pro-drop language is an inadequate description, for two reasons. The first of these is that all of the Romance varieties studied here permit the null subject in certain circumstances, and strictly forbid it in others; none of them permits pro without exception, nor does any of them prohibit it uniformly throughout the syntactic and morphological environments covered in the data set. So rather than
saying that they are (or are not) pro-drop languages, it is more accurate to assert that there are instances within the languages where they do or do not act like pro-drop languages.

This leads us directly into another issue, which is related to, and implied by, the first. It concerns the lack of acknowledgement of language-internal variation in null subject usage. It is perhaps quite surprising that, despite the work done by Chomsky (1981) and Gilligan (1987) in the preceding decade, the definitions taken from works in the 1990s largely fail to address the existence of intra-linguistic null subject microvariation, instead resorting to wider, general treatments of pro-drop on a solely macroscopic, cross-linguistic level. There is no mention, either by the phrase itself or by any appropriate description pertaining to the concept, of a “mixed system” in the three quotes above. Furthermore, Radford’s strongly-worded insistence on a binary parameter is enough to conclude safely that he is not a proponent of mixed systems with regard to null subject usage.

It should be obvious at this point that our data are not compatible with any classification system that forces us to adopt either a positive or a negative “pro-drop parameter” setting for the languages under investigation. We can no longer justify assigning languages a [+pro-drop] designation because we have identified situations where they do not act accordingly. If we recall that Catalan is the “most pro-drop” of our six Romance varieties, as characterised in our language hierarchy from Section 4.5, and that the relatively small number of syntactic and morphological conditions tested in our data do not begin to cover the myriad grammatical environments that could affect null subject usage, it becomes increasingly clear how tenuous an assertion it would be to claim that Catalan, taken as one indivisible entity, is [+pro-drop] when we have identified five instances, in just a limited range of linguistic conditions tested, where it does not behave as such. We can imagine therefore how even less veracious it would be to designate the languages lower than Catalan on our null subject hierarchy as having an unconditionally positive pro-drop parameter setting.

We thus maintain that any framework incorporating a single pro-drop parameter that is applied to a language as a whole is not workable. Therefore, if we still want to attempt a tenable reconciliation of our data with a PPT model, it will be necessary to make serious modifications to the schema, and even this action will not guarantee the elimination of all of the problems highlighted above.
5.1.2 Attempting to fit the data into a GB model

5.1.2.1 Setting up a hypothetical framework

In order to accommodate the observed patterns in null subject usage for our six Romance varieties, it would be necessary to replace the all-encompassing “pro-drop parameter” with a number of sub-parameters (or viewed differently, to break down the larger parameter into its smaller constituent components). The minimum number of sub-parameters required to account for all of our data would be seven, one for each of the three features that can be indicated in the predicate, one for each of the three syntactic environments in which we collected our data, and one to indicate whether the subject is co-referent with the topic antecedent. So the seven sub-parameters would be:

[+/- gender] – positively set if gender-indicating morphology is present in the predicate and uniquely identifies the intended antecedent

[+/- number] – positively set if number-indicating morphology is present in the predicate and uniquely identifies the intended antecedent

[+/- person] – positively set if person-indicating morphology is present in the predicate and uniquely identifies the intended antecedent

[+/- T, MC] – positively set if the construction consists of a topic followed by main clause only

[+/- T, MC, SC] – positively set if the construction consists of a topic, followed by main clause, which in turn is followed by a subordinate clause

[+/- 2S] – positively set if the construction consists of two sentences

[+/- topic] – positively set if the subject pronoun is co-referent with a topic antecedent

These seven sub-parameters are each indispensable. We cannot explain all of the microvariation witnessed in our data with any fewer, without complicating the

---

4 It should be noted that by the phrase “uniquely identifies”, we include both direct indication of intended antecedent (e.g., verbal desinence uniquely specifies the co-referent of the subject pronoun) and indirect indication (e.g., clitic object pronoun identifies the antecedent not intended to be co-referenced with the subject pronoun, thereby leading to, by default and the process of elimination, the identification of the antecedent that is the intended one).
sub-parameters to the point of their being largely *ad hoc* and useless in general contexts. To illustrate the necessity of all seven components of this hypothesised framework, we list below minimally different items from our data that result in opposite licensing of the null subject, and that can thus only be accounted for by one of these seven sub-parameters:

\[
\begin{align*}
[+ \text{ gender}] & \quad (28cs) \quad \text{Juan}_1 \text{ va a quedar con Maria}_2 \text{ hoy}. \text{ (pro}_2, \text{ Le}_1 \text{ odia.} \\
[- \text{ gender}] & \quad (26cs) \quad \text{Juan}_1 \text{ va a quedar con Miguel}_2 \text{ hoy}. \text{ *pro}_2 / \text{ Él}_2 \text{ le}_1 \text{ odia.} \\
[+ \text{ number}] & \quad (30ep) \quad \text{O Joao}_1 \text{ encontra-se com os vizinhos}_2 \text{ hoje}. \text{ (pro}_2, \text{ Odeiam-no}_1. \\
[- \text{ number}] & \quad (26ep) \quad \text{O Joao}_1 \text{ encontra-se com o Miguel}_2 \text{ hoje}. \text{ *pro}_2 / \text{ Ele}_2 \text{ odeia-o}_1. \\
[+ \text{ person}] & \quad (32bp) \quad \text{Encontro-me com o meu pai}_2 \text{ hoje}. \text{ (pro}_2, \text{ Odiava-me}_1. \\
[- \text{ person}] & \quad (26bp) \quad \text{O Joao}_1 \text{ se encontra com o Miguel}_2 \text{ hoje}. \text{ *pro}_2 / \text{ Ele}_2 \text{ odeia ele}_1. \\
[+ \text{ T, MC}] & \quad (6ct) \quad \text{Els veins}_2 \text{ d'en Joan}_1, \text{ *pro}_1 / \text{ ell}_1 \text{ els}_2 \text{ menspreua.} \\
[- \text{ T, MC}] & \quad (22ct) \quad \text{Els veins}_2 \text{ d'en Joan}_1, \text{ sabem que (pro}_1 \text{) els}_2 \text{ menspreua.} \\
[+ \text{ T, MC, SC}] & \quad (22ct) \quad \text{Els veins}_2 \text{ d'en Joan}_1, \text{ sabem que (pro}_1 \text{) els}_2 \text{ menspreua.} \\
[- \text{ T, MC, SC}] & \quad (6ct) \quad \text{Els veins}_2 \text{ d'en Joan}_1, \text{ *pro}_1 / \text{ ell}_1 \text{ els}_2 \text{ menspreua.} \\
[+ \text{ topic}] & \quad (29it) \quad \text{Giannii, incontra i vicini}_2 \text{ oggi}. \text{ (pro}_1 \text{) Li}_2 \text{ odia.} \\
[- \text{ topic}] & \quad (6it) \quad \text{I vicini}_2 \text{ di Gianni}_1, \text{ *pro}_1 / \text{ lui}_1 \text{ li}_2 \text{ odia.} \\
[+ \text{ 2S}] & \quad (1pr) \quad \text{El padre}_2 \text{ de Juan}_1, \text{ (pro}_1 \text{) lo}_2 \text{ odia.} \\
[- \text{ 2S}] & \quad (2pr) \quad \text{El padre}_2 \text{ de Juan}_1, \text{ *pro}_1 / \text{ él}_1 \text{ lo}_2 \text{ odia.} 
\end{align*}
\]

The nature of these seven sub-parameters varies as well. The three featural sub-parameters, [+/- gender], [+/- number], and [+/- person], are mutually independent from each other. The way one of these sub-parameters is set has no bearing on how the other two are set. Therefore, these three binary sub-parameters may be set in \(2^3\), or eight, different combinations.

The three sub-parameters specifying for syntactic environment do not behave like their featural counterparts; they interact without independence from each other. This is because no item of data can be set positively for more than one of the three
sub-parameters simultaneously. A sequence from our data can either be [+ T, MC], or it can be [+ T, MC, SC], or it can be [+ 2S]. It cannot be, for instance, both [+ T, MC] and [+2S] at the same time. So one, and only one, of these sub-parameters must be set positively for the purposes of our data from Chapter 4. That entails, additionally, that the three syntactic sub-parameters cannot all be set negatively at any one time. In the absence of any other syntactic constructions under investigation in the present research, having the sub-parameter setting combination { [- T, MC], [- T, MC, SC], [- 2S] } would imply a null or degenerate syntactic environment that cannot be evaluated with any relevance in this framework. Each of these three syntactic sub-parameters, however, can interact independently with each of the three featural sub-parameters described above.

The sub-parameter specifying topichood acts independently of the other six, since two options are always available, irrespective of which syntactic environment is being considered and which morphological properties are present. We have the choice throughout of co-reference with the topic antecedent, or co-reference with the non-topic antecedent. Therefore, the total number of different combinations of sub-parameter settings we have for the microcosmic null subject system in the present work is [8 combinations of gender, number, and person settings] X [3 choices for syntactic environment] X [2 options for co-reference], or 48 total possibilities. This number may come as a staggering surprise, considering both the paucity of data used in the present research (it does not come close to reflecting all the possible syntactic conditions and morphological expressions that exist in languages), and also the fact that the traditional method of explaining null subjecthood in the GB model consisted of just the one "pro-drop parameter", with two possible settings. However, we feel justified in claiming that, if the data had to be fitted into this syntactic framework, the 48 combinations offered by our seven sub-parameters would be a much more accurate representation of the nature of null subject licensing than the common simplistic model offered by most textbooks on syntax and grammar.

5.1.2.2 The proposed model at work

The application of these seven sub-parameters would most likely operate as follows. For any particular combination of the proposed seven sub-parameters in a
language, we witness either the ability of the subject pronoun to be dropped, or the requirement that it be overt. To illustrate this, consider the following Castilian datum:

(1) [26cs]\(^5\) \(Juan_1\) va a quedar con Miguel_2 hoy. *pro_2 / Él_2 le_1 odia.

'John is going to meet Michael today. He (= Michael) hates him (= John).'

We see in this sequence of two sentences the following properties, and the corresponding settings for the associated sub-parameters:

- gender-indicating morphology cannot distinguish between the two potential antecedents Juan and Miguel, both of whom are morphosyntactically possible co-referents with the anique masculine direct object pronoun le in the second sentence, and would also show subject-verb agreement with odia. Therefore, this sub-parameter is set as [- gender].

- number-indicating morphology cannot distinguish between the two potential antecedents, since both of them are singular, as is the direct object pronoun. Both are compatible with the verb's specification for a singular subject. This sub-parameter is set as [- number].

- person-indicating morphology cannot distinguish between the two potential antecedents. Juan, Miguel, and le all carry third-person features, as does the verbal desinence on odia. This sub-parameter is set as [- person].

- the syntactic environment under consideration is a sequence of two sentences, so the three syntactic sub-parameters are set as [- T, MC], [- T, MC, SC], and [+ 2S ].

- the subject of the second sentence is co-referenced in (1) with the non-topic antecedent, so [- topic ] would be the appropriate setting for this sub-parameter.

Therefore, for this combination of settings, we observe that the subject pronoun is obligatorily overt. Thus, for any situation in Castilian with the sub-parameters set as they are in (1), Castilian would be predicted to act like a [-pro-drop] language. This can be clearly summarised for our current example in tabular form:

\(^5\) This notation indicates that (1) is the equivalent of (26cs) in Chapter 4.
To demonstrate how Castilian microvariation could be accounted for under this hypothetical framework, we would need to apply the same procedure onto all the other items of Castilian data, and compare these results with what we discovered from our analysis of (1) above. As an illustration, we do this here for item (13cs), which shows slight differences in its evaluation from what was conducted for (1):

(2) [13cs]  

Los vecinos₁ de todas nosotras₂, (pro₁) nos₂ odian.  

‘Our (fem.) neighbours, they hate us.’

The featural sub-parameter settings are [-gender]⁶, [-number], and [+person]. (2) is an example of a topic followed by main clause, so the three syntactic sub-parameters would be set as [ + T, MC ], [ - T, MC, SC ], and [ - 2S ]. Finally, the subject is co-referenced with the topic, and hence earns the [ + topic ] designation. So for a morphosyntactic environment typified by (2), the grouping of sub-parameters would be as follows:

[gender] [number] [person] [T, MC] [T, MC, SC] [2S] [topic]  [pro-drop]  

- - + + - + +

This results, then, in the designation of Castilian as a [+pro-drop] language, for this specific combination of linguistic circumstances. (1) and (2) can be compared more directly by including their respective settings in the same table; this would provide an even clearer treatment of monolingual microvariation, as seen below:

---

⁶ Although the two antecedents differ in their gender feature, neither nos nor odian in the predicate is able to reflect this, since both are gender-neutral. This accounts for the [-gender] setting.
Extrapolating the procedure above to all the other pieces of Castilian data would provide the complete version of our grid above. This task could then be extended to the other Romance varieties under consideration. Once this has been accomplished, we would have a thorough account of our data in a GB framework. Further combinatorial analysis could be attempted to find out mathematically how these sub-parameters correlate with each other and with the final [pro-drop] setting, both intra-linguistically and cross-linguistically, but this is beyond the scope of the current exposition.

5.1.2.3 How to analyse and compare in this model

According to the traditional approach to pro-drop in the GB framework, languages were contrasted in a rudimentary fashion. Languages were apparently either [+pro-drop], such as Spanish and Italian, or [-pro-drop], such as English and French. Under this multiple sub-parameter model presently being discussed, the basic premise of cross-linguistic comparison is the same, except that comparisons of languages are made in view of the settings of seven sub-parameters instead of just the single broad parameter.

For example, consider the six analogous versions in our data of a particular linguistic situation (two-sentence sequence; the two antecedents differing in gender features, but not in number and person features; and co-reference of subject with the non-topic antecedent):

(3a) [28ct]  *En Joan, es reuneix amb en Maria, avui. (pro₂) El₁, menyspreua.
(3b) [28cs]  Juan₁, va a quedar con Maria₂ hoy. (pro₂) Le₁, odia.
(3c) [28ep]  O João₁, encontra-se com o Maria₂ hoje. *pro₂ / Ela₂, odeia-o₁.
We point out from the data above that only Catalan and Castilian allow the null subject at the beginning of the second sentence. In Italian, Puerto Rican Spanish, and both varieties of Portuguese, an overt subject pronoun is compulsory to avoid ungrammaticality.

The corresponding sub-parameter settings for all six versions of (3) are identical, of course, since all six sequences are language-specific versions of the same linguistic environment under review; namely, \([ + \text{gender}], [ - \text{number}], [ - \text{person}], [ -\ T,\ MC], [ -T,\ MC,\ SC], [ + 2S ],\) and \([ -\ \text{topic}].\) Therefore, cross-linguistic comparisons of null subject usage within a GB framework would have to be done with respect to any one fixed combination of sub-parameter settings. In other words, we would have to consider first the combination of settings \([ + \text{gender}],[ - \text{number}],\ [ -\ \text{person}], [ -T,\ MC],\ [ -T,\ MC,\ SC],\ [ + 2S ],\ [ -\ \text{topic}].\) and then, for that particular setting, we could describe the variation in null subject usage of the languages examined, as follows:

\[
[ + \text{gender}], [ - \text{number}], [ - \text{person}], [- T, MC], [ - T, MC, SC], [ + 2S ], [ - topic ]
\]

<table>
<thead>
<tr>
<th>Language</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Castilian</td>
<td>(+ pro-drop)</td>
</tr>
<tr>
<td>Catalan</td>
<td>(+ pro-drop)</td>
</tr>
<tr>
<td>European Portuguese</td>
<td>(- pro-drop)</td>
</tr>
<tr>
<td>Italian</td>
<td>(- pro-drop)</td>
</tr>
<tr>
<td>Brazilian Portuguese</td>
<td>(- pro-drop)</td>
</tr>
<tr>
<td>Puerto Rican Spanish</td>
<td>(- pro-drop)</td>
</tr>
</tbody>
</table>

It would not be possible to compare the licensing of \textit{pro-drop} among these languages in any other fashion. We certainly could not attempt the above exercise
with respect to just one sub-parameter setting, for two reasons. First, we have already argued that, generally speaking, each language shows significant internal variation with regard to even just one of the sub-parameters. Second, the interaction of sub-parameters must be taken into account as well, so any approach not incorporating the combination of all the relevant sub-parameters would merely lead to the formulation of imprecise, sweeping statements.

It becomes increasingly apparent how impractical this approach is starting to become. Not only are we faced with an overwhelming, if not infinite, number of sub-parameters to take into account, but we must also consider all possible combinations of settings for the countless sub-parameters, an exhausting task the benefits of which have not yet been conclusively determined. Furthermore, this hypothetical model lacks lucidity, which is undesirable but bearable if the framework is otherwise sound. However, since the flawlessness of this proposal has not been established, the lack of clarity is arguably a fatal error. We can thus understand one of the fundamental problems of accounting for the null subject phenomenon within a GB framework: the traditional approach of a language-wide positive or negative pro-drop parameter setting is intuitively graspable but oversimplistic and largely inaccurate. The other extreme of this is a model that, while being a much better reflection of the actual situation in null subject usage, is extremely complicated, not as easily understandable, and may still prove to be an inefficient and sub-optimal method of viewing the pro-drop phenomenon. Any attempt to account for the null subject in a GB framework is thus working at a significant handicap. We explore later just how disadvantaged it is in comparison to other suggested models.

5.1.2.4 The possibility of feasible substitute models

We should mention here as well that there are alternative methods to constructing a revised null subject model in the GB framework. The hypothetical version that we have just explored is not the only one possible, but it does accurately indicate the types of issues and pitfalls that any such proposal would have to address satisfactorily in order to be successful. One alternative to our proposal would be to conflate the three syntactic sub-parameters [T, MC], [T, MC, SC], and [2S] into one ternary sub-parameter, with three settings to represent each one of the three syntactic
environments being investigated in our work. There would be a slight economical improvement to our model in the reduction of the number of sub-parameters from seven to five, but this would not change the fact that 48 total combinations of sub-parameter settings are possible in our framework, nor would it alter the theoretical arguments in favour of and against the model developed in this section. However, any benefits from this structural reduction would be offset by the introduction of a ternary sub-parameter. We are disinclined to lobby in favour of non-binary parameters because of the ripple effect. That is, the interdependence of various language modules constituting a grammar means that a switch in one parameter has measurable effects on many other aspects of the linguistic system. By limiting the variation in a parameter to two options, we are better able to control how the grammar is affected when one parameter switches settings. However, if we open the possibility that a parameter may shift to more than one alternative setting, then we must be prepared for an exponential explosion in the number of changes resultant in the language, and the number of different possible grammars this could produce. This would have serious consequences on language acquisition and efficiency of representation. Only when the option of a binary parameter is strictly unavailable should we consider a parameter with multiple settings. Otherwise, we argue that the best approach is to keep the parameters binary and work around their imperfections through other methods.

Another possibility would be to incorporate topichood into the syntactic sub-parameters, so that we understand [+2S] to mean subject co-referent with topic antecedent in a two-sentence sequence, and [-2S] to mean subject is co-referent with the non-topic antecedent. This would eliminate a separate value for topic, and reduce our inventory to six sub-parameters. However, 48 combinations would still be possible, eight for the various combinations from the feature sub-parameters, multiplied by the six values from the syntactic sub-parameters (([-+2S], [+2S], [+2S], [-2S], [-T, MC], [-T, MC], [+T, MC, SC], and [-T, MC, SC]). It should be noted here as well that only one of the three syntactic sub-parameters could ever be set positively at any one instance. That is, it would be impossible for two or more of these sub-parameters to carry a positive setting simultaneously. For example, [+2S] would necessarily preclude the [+T, MC, SC] and the [+T, MC] settings because otherwise this particular combination of settings would imply a situation where the syntactic environment under consideration is both a two-sentence structure and a topic, main clause, followed by subordinate clause structure. Based solely on the data under
consideration in this work, this is an impossibility. Each of our examples from Chapter 4 can be classified only as either a 2S structure, or a T, MC, SC structure, or a T, MC structure, mutually exclusive of one another.

5.1.2.5 Using parameters to explain a non-parametric phenomenon?

To summarise then, in the hypothetical versions of the null subject “parameter” in a GB framework presented here, we would no longer say that Catalan is a [+pro-drop] language and that English is a [-pro-drop] language. Instead, we would have to consider combinations of sub-parameters, and then say that, under any specific combination of sub-parameter settings, a language is [+pro-drop] or [-pro-drop]. Characterisations such as [+pro-drop] or [-pro-drop] would therefore not be applicable to any language as a whole, but rather to each language under certain morphosyntactic circumstances represented by the settings of the various sub-parameters.

Perhaps it may be stating the obvious, but it is becoming gradually clearer that the principle reason why null subject usage is not adequately accounted for in a PPT model, either intra-linguistically or cross-linguistically, is that this phenomenon is not a parameter to begin with. Since pro-drop cannot be satisfactorily described through parameters, we should not attempt a forced squeezing of our data into a PPT framework then. Variation in the licensing of pro is best described as a cline or continuum, as per Ross’ (1982) adaptation of McLuhan’s (1964) “hot, warm, cool, and cold” language metaphor (Huang 1984: 531). It is a continuous ramp; therefore, we need to find a model that is not predicated upon discrete yes-or-no distinctions or rigid black-or-white categorisations, but instead allows for smooth and gradual variation. With that goal in mind, we now turn to an investigation of the Minimalist framework in an attempt to find such a model.
5.2 Limitations of the Minimalist framework

Here we inspect another framework within PPT, the Minimalist Program (henceforth MP), to determine whether this provides a better model for our findings.

5.2.1 EPP and feature strength

A suitable starting point would be Chomsky (1995), generally credited as the germinal work outlining this minimalist approach to linguistic theory. Chomsky does not specifically tackle pro-drop in this work per se. Instead, we have to infer his proposal for how MP would address the null subject phenomenon from his expositions of both the Extended Projection Principle (henceforth EPP) and the notion of feature strength within this framework. It is only through extrapolation of these developed hypotheses and their subsequent applications as relevant to the licensing of pro that we are able to understand how Chomsky intends to treat pro-drop within MP.

Initial mention of the possible existence of a phonetically null subject is found in Chomsky’s minimalist formulation of the EPP. He speculates that “[t]he Extended Projection Principle, which requires that [Spec, IP] be realized (perhaps by an empty category), reduces to a morphological property of T: strong or weak NP-features” (Chomsky 1995: 199). The closest that Chomsky comes to discussing the concept of null subjecthood occurs in his presentation on feature strength, where he writes that “a formal feature may or may not be strong, forcing overt movement that violates Procrastinate. [...] Thus, the Extended Projection Principle (EPP) plausibly reduces to a strong D-feature of I” (1995: 232). These statements must be considered in conjunction with Chomsky’s proposals on feature checking. A strong feature, Chomsky notes, “triggers an overt operation, before Spell-Out. [...] It thus triggers a rule that eliminates it: [strength] is associated with a pair of operations, one that introduces it into the derivation (actually, a combination of Select and Merge), a second that (quickly) eliminates it. [...] We also virtually derive the conclusion that a strong feature triggers an overt operation to eliminate it by checking” (1995: 233).

In order for pro-drop to be fitted into this alternate version of the PPT model

7 Or, more specifically, T, under the Split-INFL hypothesis as initially stipulated by Pollock (1989).
of grammar, Chomsky’s explanations above, taken all together, would likely have to be interpreted as follows. We would need to postulate that an uninterpretable strong D-feature on T would have to be checked via overt movement of a DP to [Spec, TP] before Spell-Out. The crucial difference lies in the nature of the unit moved to check this strong D-feature. In Italian, this can be the phonetically realised pro or a lexical DP, whereas in English, a lexical DP is the only option. More broadly speaking, in what are traditionally called null subject languages, pro (alongside overt forms) is featured in the inventory of pronominal elements stored in the lexicon, whereas in non-null subject languages, it is not. Therefore, in these languages pro is not available to check this strong D-feature on T. Variation is hence rooted in the lexica.

Holmberg (2002) proposes a deletion analysis of pro-drop. He suggests that the numeration begins with an overt pronoun, but the specific D-related properties of T in null subject languages allow these pronominals that are moved to [Spec, TP] to be deleted. A variant of this proposal is offered by Roberts (2004), who argues that the D-features on T and on the pronoun are identical in null subject languages, thus licensing deletion of the pronominal. On the other hand, in non-null subject languages, T does not carry the same D-feature specifications, therefore prohibiting the deletion of the pronominal.

There exist a number of slightly different Minimalist accounts explaining null subjecthood from the ones above. One such alternative to moving DP to [Spec, TP] involves the raising of finite verbs to T. Chomsky writes:

"[...] suppose that T has a strong V-feature and a strong nominal feature (person, we have assumed; D or N in categorial systems). It has always been taken for granted that the strong V-feature is satisfied by V-raising to T (French vs English), not VP-raising to [Spec, T], and that the strong nominal feature is satisfied by raising of the nominal to [Spec, T] (EPP), not raising of its head to T. But the theoretical apparatus provides no obvious basis for this choice."

(Chomsky (2001: 37-38), reproduced from (1999: 31))

So the message presented here is that there is nothing in the Minimalist framework that strictly requires T to satisfy its D-feature by XP-movement (i.e., raising the subject to [Spec, TP]). It would be just as legitimate for this checking to be conducted through Move/Merge X (e.g., V-to-T raising) instead of Move/Merge XP. This is consistent with the proposal offered by Alexiadou and
Anagnostopoulou (1998). They maintain that pro-drop languages have [+D] verbal agreement, that this feature is strong, that the requirement to check this strong D feature of AGRSP (i.e., the EPP) causes V-raising in null subject languages, and that the reason why the EPP is checked by the verb and not by the subject NP is due to considerations of economy (Alexiadou and Anagnostopoulou 1998: 518-521). So, for instance, in Spanish and Greek, T has a pronominal feature, and the EPP is satisfied through overt V-to-T movement. In languages like English, T has no such pronominal feature, and so the EPP is satisfied only through the raising of an overt DP to [Spec, TP].

Alexiadou and Anagnostopoulou hypothesise that the raising of finite verbs to T serves two purposes: it simultaneously satisfies the strong V-feature as well as the strong D-feature on T. Therefore, they claim that there is no role for lexical subject DPs or pro in fulfilling the EPP requirement, because the D-feature on T has already been checked by V-movement. Alexiadou and Anagnostopoulou suggest that the "AGR affix replaces DP in its EPP-licensing capacity" (1998: 531), and combined with their argument that [Spec, TP] as a subject position is not licensed in Greek and Spanish8 (1998: 516), they speculate as to whether pro does in fact have to exist. While they do not commit fully to a pro-less syntactic framework, they do mention its possibility and that this may become clearer with additional research (Alexiadou and Anagnostopoulou 1998: 531-533). We agree with the authors and do not attempt to address the issue here, leaving it for future study9.

One hindrance to Alexiadou and Anagnostopoulou’s approach is that there is an assumption that in all pro-drop languages, verbs raise from within the VP to the T/Agr projection (i.e., high raising). However, Ledgeway and Lombardi (2005 forthcoming) show that there are many pro-drop dialects of southern Italy where the finite verb raises only to low positions, as in English. Therefore, in order to make

8 Alexiadou and Anagnostopoulou, citing Chomsky (1995) claim that “specifier positions are only projected when a strong nominal feature forces merging or movement of an XP to the functional category carrying this feature; thus, it follows that if a language lacks Move/Merge XP to check the EPP, then in this language Spec, AGRS is not projected” (1998: 501). Jonas and Bobaljik (1993) and Bobaljik and Jonas (1996) mention in their discussion of the parametric availability of [Spec, TP] that this position is not licensed for subjects in Romance.

9 Good starting points to investigate this debate in depth include Pollock (1997), Nash and Rouveret (1997), Manzini and Roussou (1999), and Manzini and Savoia (2002).
these discoveries consistent with their proposal, Alexiadou and Anagnostopoulou would have to separate verb movement from the EPP, because otherwise we would expect these southern Italian dialects not to be null subject languages.

Irrespective of the manner in which we attempt to account for the D-feature in the T domain, be that through move/merge XP or move/merge $X^0$, it is still the case that we have difficulties in reconciling our data with any version of MP because of its very nature. In particular, a Minimalist treatment of null subjecthood does not appear to accommodate intra-linguistic variation to the extent desired. This mirrors the primary drawback of the traditional GB account of a language-wide “pro-drop parameter”. We repeat here that it is no longer feasible to assert that languages, as a whole, are either able to drop subject pronouns, or are unable to drop them. We explored previously how, if we were forced to fit our data into a GB model, we found it necessary to deconstruct our traditional pro-drop parameter into a series of relevant, constituent sub-parameters in order to account for null subject intra-linguistic variation.

A parallel amendment to the MP proposals above, taking into consideration specific morphosyntactic environments, would therefore be in order. In light of the resemblance between the “pro-drop parameter” and the binary options dictating how the EPP is satisfied in MP, it is reasonable to argue that the pitfalls identified with explaining null subjecthood through a GB framework apply to a large extent to any MP account as well. If we had to accommodate our data into a Minimalist model, we would soon notice the following significant risk. By proposing a parallel formulation to that which was offered for the GB model in Section 5.1.2, we would have to say that the availability of pro to check the strong D-feature on T is dependent upon certain circumstances, such as the syntactic environment under consideration, which features (understood here in the sense of “person”, “number”, and “gender”, not D-feature or T-feature) distinguish between the two antecedents in question, and whether co-reference is with a previous topic. We consequently end up with an elaborate and complicated system, the benefits of which would be inconclusive at the present moment. Furthermore, as in the situation with the multiple sub-parameters, it is not certain whether the vast number of individual factors that combine to determine what can check the strong D-feature on T in a particular linguistic situation would be finite. Even if it were, it is highly debatable whether this model would be consistent with our
current knowledge of language acquisition and the efficient representations of mental grammars. Aside from that, we would also have to address whether this proposal is merely ad hoc, descriptive, and circular, thus lacking in any explanatory value, and whether we are forcing a parameterised account onto a linguistic phenomenon that should not be treated in such a manner.

5.2.2 Attempting to fit the data into a MP model

One possible formulation of our data into a Minimalist framework may be constructed as follows. We notice that the different Romance varieties under consideration here display a full range of sensitivities to the presence of antecedent-distinguishing features in determining whether pro is licensed. That is, on one end we have one Romance variety, Puerto Rican Spanish, that only looks for whether the proposed null subject pronoun is co-referent with the previous topic. If it is, pro is licensed. There is no possibility for a non-topic to serve as the antecedent to pro, no matter how closely their features match. Located at the other extreme of the spectrum is Catalan, a language that permits co-reference between a null subject and any antecedent, regardless of its topichood status, so long as said antecedent perfectly matches pro with respect to all of its characterising features. Positioned in-between the two poles represented here by Puerto Rican Spanish and Catalan are our other Romance varieties, which all tolerate co-reference between pro and any previous topical antecedent if their features match. Where they differ, however, is the extent to which they tolerate co-reference between pro and a non-topic antecedent, when the respective features for the two entities match. In particular, more pro-friendly varieties such as Castilian will look at multiple features before calling for an overt subject pronoun to be spelled out, whereas those varieties more resistant to pro, such as Brazilian Portuguese, are less “patient” and will only bother to look at a fewer number of features before demanding an overt subject pronoun.

What this hypothesis signifies is the acknowledgement that each language has a different depth into which it is willing to look at the feature matrix before making its decision to license pro. To illustrate this, we consider our data from Chapter 4 in this slightly alternative manner. Let us first examine Puerto Rican Spanish. The following sentences will suffice to show that the only factor pertinent to the tolerance
of pro in this variety is the topichood of the intended co-referenced antecedent:

- co-reference between subject and topic antecedent

(4) [25pr] \(Juan_1 \text{ va a encontrarse con Miguel}_2 \text{ hoy. (pro}_1 \text{) Lo}_2 \text{ odia.}\)

- co-reference between subject and non-topic antecedent, where antecedents are not semantically discernible by gender, number, or person

(5) [26pr] \(Juan_1 \text{ va a encontrarse con Miguel}_2 \text{ hoy. *pro}_2 / \text{Él}_2 \text{ lo}_1 \text{ odia.}\)

- co-reference between subject and non-topic antecedent, where antecedents are semantically discernible by gender

(6) [28pr] \(Juan_1 \text{ va a encontrarse con Maria}_2 \text{ hoy. *pro}_2 / \text{Ella}_2 \text{ lo}_1 \text{ odia.}\)

- co-reference between subject and non-topic antecedent, where antecedents are semantically discernible by number

(7) [30pr] \(Juan_1 \text{ va a encontrarse con los vecinos}_2 \text{ hoy. *pro}_2 / \text{Ellos}_2 \text{ lo}_1 \text{ odian.}\)

- co-reference between subject and non-topic antecedent, where antecedents are semantically discernible by person

(8) [32pr] \((pro)_1 \text{ Voy a encontrarme con mi padre}_2 \text{ hoy. *pro}_2 / \text{Él}_2 \text{ me}_1 \text{ odiaba.}\)

Any feature that differentiates between the two competing antecedents, whether that be person, number, or gender, is irrelevant in Puerto Rican Spanish; the only crucial determination is whether the antecedent intended for co-referencing is a previous topic. So we may assert that topichood is the only contributing factor, among those under consideration in the present research, in the licensing of Puerto Rican Spanish pro.

Next we consider the Romance variety that is one step above Puerto Rican Spanish on our proposed hierarchy, Brazilian Portuguese. Parallel constructions are presented below:
• co-reference between subject and topic antecedent

(9) [25bp]  O João₁ se encontra com o Miguel₂ hoje. (pro₁) Odeia ele₂.

• co-reference between subject and non-topic antecedent, where antecedents are not semantically discernible by gender, number, or person

(10) [26bp]  O João₁ se encontra com o Miguel₂ hoje. *pro₂ / Ele₂ odeia ele₁.

• co-reference between subject and non-topic antecedent, where antecedents are semantically discernible by gender

(11) [28bp]  O João₁ se encontra com a Maria₂ hoje. *pro₂ / Ela₂ odeia ele₁.

• co-reference between subject and non-topic antecedent, where antecedents are semantically discernible by number

(12) [30bp]  O João₁ se encontra com os vizinhos₂ hoje. *pro₂ / Eles₂ odeiam ele₁.

• co-reference between subject and non-topic antecedent, where antecedents are semantically discernible by person

(13) [32bp]  (pro₁) Encontro-me com o meu pai₂ hoje. (pro₂) Odiava-me₁.

The difference we note here from the previous situation is that the only qualities relevant to the licensing of pro are topichood and whether person can distinguish between the two competing antecedents. It appears as if number and gender differences in instances of switch references do not enter into the deliberation to license pro. We may therefore pinpoint where Puerto Rican Spanish and Brazilian Portuguese diverge by stating that in the former, the factor that affects the tolerance of pro is \{topic\}, whereas the subset of factors that affect the same in the latter consists of \{topic, person\}. Analyses of the remaining Romance varieties under consideration in this work would proceed in the same fashion. So for Italian and European Portuguese, the relevant subset of factors would be \{topic, person, number\}, and for Castilian, \{topic, person, number, gender\}. Catalan allows co-referencing between any two entities with identical features, so it does not look into the internal structure of the feature matrix itself.

The specific details of this explanation above only work, of course, with the
• co-reference between subject and topic antecedent

(9) [25bp]  O João₁ se encontra com o Miguel₂ hoje. (pro₁) Odeia ele₂.

• co-reference between subject and non-topic antecedent, where antecedents are not semantically discernible by gender, number, or person

(10) [26bp]  O João₁ se encontra com o Miguel₂ hoje. *pro₂ / Ele₂ odeia ele₁.

• co-reference between subject and non-topic antecedent, where antecedents are semantically discernible by gender

(11) [28bp]  O João₁ se encontra com a Maria₂ hoje. *pro₂ / Ela₂ odeia ele₁.

• co-reference between subject and non-topic antecedent, where antecedents are semantically discernible by number

(12) [30bp]  O João₁ se encontra com os vizinhos₂ hoje. *pro₂ / Eles₂ odeiam ele₁.

• co-reference between subject and non-topic antecedent, where antecedents are semantically discernible by person

(13) [32bp]  (pro₁) Encontro-me com o meu pai₂ hoje. (pro₂) Odiava-me₁.

The difference we note here from the previous situation is that the only qualities relevant to the licensing of pro are topichood and whether person can distinguish between the two competing antecedents. It appears as if number and gender differences in instances of switch references do not enter into the deliberation to license pro. We may therefore pinpoint where Puerto Rican Spanish and Brazilian Portuguese diverge by stating that in the former, the factor that affects the tolerance of pro is {topic}, whereas the subset of factors that affect the same in the latter consists of {topic, person}. Analyses of the remaining Romance varieties under consideration in this work would proceed in the same fashion. So for Italian and European Portuguese, the relevant subset of factors would be {topic, person, number}, and for Castilian, {topic, person, number, gender}. Catalan allows co-referencing between any two entities with identical features, so it does not look into the internal structure of the feature matrix itself.

The specific details of this explanation above only work, of course, with the
two-sentence syntactic environment that we have chosen for the sake of exposition. However, similar lines of argumentation, with suitable amendments as appropriate, would be applicable to the [T, MC] and [T, MC, SC] syntactic environments we have examined in the present research; the fundamental underlying reasoning developed here would not be altered. For example, in [T, MC] constructions, where pro may not be co-referenced with the non-topic antecedent in any of our six Romance varieties, the mechanism would only look at the topichood of the antecedent. Features would be irrelevant to the licensing of pro; therefore, the subset of pertinent features in this particular situation would be \{topic\}. In [T, MC, SC] structures, Catalan, Castilian, and European Portuguese permit co-referencing between pro and the non-topic antecedent when person or number is able to distinguish between the two competing antecedents, so we may postulate that for this linguistic condition, the relevant features in those languages to the licensing of pro are \{topic, person, number\}. In Italian, Brazilian Portuguese, and Puerto Rican Spanish, on the other hand, pro may never be co-referenced with the non-topic antecedent, so the only relevant feature to the tolerance of pro is \{topic\}.

5.2.3 Looking into the feature bundles

What is most notable about the exposition above is that attention has been drawn to the internal structure of the feature matrix associated with these DPs. Languages differ as to whether they are near-sighted (i.e., they will only bother to look at a limited number of features in the feature bundle before deciding on Spell-Out) or far-sighted (i.e., they will be patient enough to consider several features within the bundle before determining whether an overt subject pronoun will be required). Puerto Rican Spanish and Brazilian Portuguese are representative of the former designation, looking not too deeply into the feature bundle before deciding on the nature of the subject pronoun. Castilian is aptly described by the latter label; it will take into consideration a higher number of features in the matrix prior to choosing an overt or covert subject pronoun. European Portuguese and Italian display intermediate tendencies in this respect. As for Catalan, the most pro-friendly of our varieties, it does not bother delving into the feature matrix at all, instead opting to accept the entire bundle as one unit, so we propose that it be designated as neither
short-sighted nor long-sighted, as these descriptions should only apply to languages that look at the internal geometry of the feature bundle.

Another point of remark is that the order of consideration of the features within the bundle is necessarily set as topic first, then person, followed by number, and finally, gender. That is to say, the subset of features \{topic, person, number, gender\} is a permutation, where the order of listing is important, rather than merely a combination that simply indicates membership in the subset. Therefore, we define our usage to signify that the designation \{topic, person, number, gender\} entails the membership as well as the order of features in the subset, and is different from, for example, \{number, topic, gender, person\}, which would imply a different order of feature considerations within the bundle. This is a crucial clarification; we hope to have convincingly shown above that there is a well-defined ordering with regard to the inspection of features within the feature bundle. Our proposal here is not tenable if we neglect to assert the necessity of feature ordering within the matrix.

We may summarise our exposition as follows. Languages examine, to different extents, the internal structure of DP feature matrices during their deliberations on whether to license pro. Our six Romance varieties, listed below in increasing order of pro tolerance, are associated with the following “featuresightedness”:

- **Puerto Rican Spanish:** \{topic\}
- **Brazilian Portuguese:** \{topic, person\}
- **Italian:** \{topic, person, number\}
- **European Portuguese:** \{topic, person, number\}
- **Castilian:** \{topic, person, number, gender\}
- **Catalan:** [looks only for any antecedent with matching features]

It is worth clarifying that the features listed above simply indicate that they are relevant to the computing mechanism; we have not yet explicitly stated how exactly they are relevant. We do so now, claiming here that the relevancy of the topic feature is simply that the computing mechanism looks for whether the intended antecedent for co-reference is a topic. The other features, person, number, and gender, indicate that the computing mechanism looks for whether our two competing antecedents may be differentiated according to those features. Thus, our feature definitions ask the following questions:
Is the intended antecedent a topic?

Does the person feature distinguish between the two competing antecedents?

Does the number feature distinguish between the two competing antecedents?

Does the gender feature distinguish between the two competing antecedents?

The mechanism is satisfied and will allow pro if it is able to generate a positive response to a feature question relevant to the language under consideration. If it has exhausted all relevant feature questions without obtaining a ‘yes’, then it must spell-out an overt subject pronoun. In the case of Catalan, which neglects to inspect within the feature bundle, instead choosing to accept it as a whole, these feature questions are never asked. The responses generated from them are impertinent; all that happens in Catalan is that feature bundles for co-referencing candidates are compared for exact matching.

To see how our proposal might operate in practice, we offer the following as a possible working of the computing mechanism. Consider the following sentence in European Portuguese:

(14) [30ep]  O João₁ encontra-se com os vizinhos₂ hoje. (pro₂)/Eles₂ odeiam-no₁.

The three pertinent DPs, ‘João’, ‘vizinhos’, and ‘eles’, display the following feature matrix specifications:

João: [+topic, +3rd person, +singular, +masculine]  
vizinhos: [-topic, +3rd person, -singular, +masculine]  
eles: [+3rd person, -singular, +masculine]

We propose that the subject of the second sentence carries no topic feature of its own, but inherits the topic feature of the antecedent with which it is co-referenced; therefore, ‘eles’ above contains only feature specifications for person, number, and gender. Alternatively, we could stipulate that for some reason, ‘eles’ either does not need to be specified for topichood (i.e., it is unvalued for this feature), or that its
setting for topichood is flexible enough to accommodate feature matching with either [+topic] or [-topic].

As for the workings of the computation, we note initially that European Portuguese is characterised as {topic, person, number}, so these three are the only features that the mechanism will consider. The first step would be to identify the intended antecedent; this would be ‘vizinhos’, because ‘João’ and ‘eles’ do not match in the number feature. Next, we look at the first feature, topic, which asks whether the intended co-reference is a topic. ‘vizinhos’ is [-topic], so the answer is no, the intended antecedent is not a topic. This forces the mechanism to consider the next relevant feature, person, which asks whether the two competing antecedents can be distinguished by this feature. Both ‘João’ and ‘vizinhos’ are [+3rd person], so the answer is again no. The mechanism must then go to the number feature, which asks whether the two competing antecedents can be distinguished by number. In this case, the answer is yes, since ‘João’ is [+singular] and ‘vizinhos’ is [+plural]. Now that a positive response has been generated, pro may be licensed, and therefore a null subject pronoun is permitted.

Now consider the computation for the following sentence and related feature bundles:

(15) [28ep]  O João₁ encontra-se com a Maria₂ hoje. *pro₂ / Ela₂ odeia-o₁.

João: [+topic, +3rd person, +singular, +masculine]
Maria: [-topic, +3rd person, +singular, -masculine]
etal: [+3rd person, +singular, -masculine]

The first steps in this computation would mirror those in the previous example up to the number feature. Whereas the two competing antecedents in (14) were distinguishable by number, thereby satisfying the mechanism and hence licensing pro, in (15) the two competing antecedents are both singular and are not distinguishable by person; a “no” answer is returned with regard to the number feature question. At this point, all three features questions relevant to European Portuguese have returned negative responses, and therefore the mechanism has not been satisfied, thus requiring an overt subject pronoun. That John and Maria differ in gender features is not
germane to the computing mechanism, because European Portuguese does not examine the gender feature question. Gender does not form part of the subset of relevant feature questions in that language, and so is never examined.

As for Catalan, the process would be simplified. Consider the following sequence:

(16) (adapted from (25ct) and (26ct) in Chapter 4)

*En Joan₁ es reuneix amb en Miquel₂ avui. (pro/ell₁₂) El₁₂ menyspreu.*

*Joan:* [+topic, +3rd person, +singular, +masculine]
*Miquel:* [-topic, +3rd person, +singular, +masculine]
*ell:* [+3rd person, +singular, +masculine]

The only action that occurs is that the computing mechanism checks for feature bundle matching between “*ell*” and any antecedent. If such an antecedent exists, then license *pro*. If more than one antecedent is eligible, as is the case here, then coreference with either antecedent is permitted, accounting for (25ct) and (26ct) in the data.

It may be helpful to summarise and formalise this proposed mechanism in the following computational flow charts below for each of our Romance varieties:

**Puerto Rican Spanish:** {topic}

1. Is subject co-referenced with previous topic?
   1a. If so, license *pro*. END
   1b. If not, require overt subject pronoun. END

**Brazilian Portuguese:** {topic, person}

1. Is subject co-referenced with previous topic?
   1a. If so, license *pro*. END
   1b. If not, look at person feature.
   2. Does person feature distinguish between competing antecedents?
      2a. If so, license *pro*. END
      2b. If not, require overt subject pronoun. END
European Portuguese and Italian: \{topic, person, number\}

1. Is subject co-referenced with previous topic?
   1a. If so, license pro. END
   1b. If not, look at person feature.

2. Does person feature distinguish between competing antecedents?
   2a. If so, license pro. END
   2b. If not, look at number feature.

3. Does number feature distinguish between competing antecedents?
   3a. If so, license pro. END
   3b. If not, require overt subject pronoun. END

Castilian: \{topic, person, number, gender\}

1. Is subject co-referenced with previous topic?
   1a. If so, license pro. END
   1b. If not, look at person feature.

2. Does person feature distinguish between competing antecedents?
   2a. If so, license pro. END
   2b. If not, look at number feature.

3. Does number feature distinguish between competing antecedents?
   3a. If so, license pro. END
   3b. If not, look at gender feature.

4. Does gender feature distinguish between competing antecedents?
   4a. If so, license pro. END
   4b. If not, require overt subject pronoun. END

Catalan: [looks only for an exact match in feature bundles]

1. Do the subject and its proposed co-referent (regardless of topichood) share the same features?
   1a. If so, license pro. END
   1b. If not, do not allow co-referencing. END

The above is one method by which we might want to incorporate our data into a Minimalist account of null subject licensing. There are, of course, a number of imperfections and ambiguities in this proposed model. It is not initially apparent, for instance, how this mechanism would interact with the feature checking operating in any Minimalist framework, and whether any interference or side effects would result. In particular, our hypothetical mechanism does not involve feature strength as
traditionally proposed in standard Minimalist literature, so it would be worthwhile to ask what ramifications the elimination of the D-feature checking requirement in the T domain would have on the rest of the linguistic system.

We have made the assumption above, without formal justification, that only the topic feature is not specified in the feature bundle of the subject of the second sentence, and that this feature is instead inherited from its co-referent antecedent or neglected altogether by the computing mechanism. It is not completely intuitive as to why the topic feature should behave or be treated in a manner different from the other features under consideration. This is surely a problem that would need rigorous inspection and correction if this framework is to be considered worthy of adoption.

Another question involves how to formalise the extent to which languages consider features in their deliberations on the licensing of pro. We have suggested here that languages higher up on our pro-drop hierarchy are farer-sighted (or more patient) when it comes to the feature matching mechanism because they take into account more features before deciding on the nature of the subject pronoun, whereas languages lower on the scale are nearer-sighted (or less patient) because they consider fewer features in the process. Additionally, we would need to devise a way to order the features within the bundle, and elaborate upon how the mechanism would observe this order in its computation. We would have to find methods of incorporating specific instructions of these operations into a Minimalist framework.

These are just but a few of the myriad issues that would need resolving before we are to accept this version of the Minimalist framework presented here. Regardless of the strengths and weaknesses of this potential model, the exposition above reinforces the observation that little has been done in Minimalism on the internal structure of the feature matrix, much less on the specific ordering of the constituents within the matrix itself. Radford does not make any reference to either of these issues in his section on feature bundles (1997: 62), nor, more recently, does Adger (2003: 52) in his introduction to Minimalist syntax; these two syntacticians are not alone in failing to explore the possibility that feature bundles are internally complex structures. What our hypothetical Minimalist framework demonstrates, however, is that there is evidence that suggests that it may indeed be necessary for the linguistic computing mechanism to look within the feature bundles, instead of just accepting them as

---

10 Biberauer (2003: 103ff.), with particular reference to Afrikaans, is one of the few works that does examine the internal geometry of feature bundles.
indivisible, packaged entities.

Our examination of the two versions of Chomsky's PPT framework highlights the fact that, based on our data from Chapter 4, null subject behaviour is best described through clines, both in the relative rankings of the languages against each other and also in the comparative relevance of the various grammatical and syntactic factors at work. The continua inherently represented in our characterisations of null subject usage, whether it be in the pro-friendly nature of the languages under study, the varying importance of our features, or the extent to which languages look within DP feature bundles, suggest that pro-drop should be conceptualised as a hierarchy involving the interactions of discrete, pertinent factors, and that this model may be most compatible with Optimality Theory. It is to this task that we now direct our attention.
Chapter 6: Appealing to Optimality Theory

We now attempt to show that accommodating our data from Chapter 4 into an Optimality Theoretic (henceforth OT) framework is a better solution than trying to explain it using traditional Chomskyan syntactic models. This will be pursued by briefly reiterating the faults of applying PPT models to our data, and demonstrating how many of these shortcomings are satisfactorily resolved once we adopt this alternate OT approach.

6.1 Advantages over Principles and Parameters

6.1.1 Getting rid of parameters

It was highlighted in Sections 5.1.2 and 5.2.2 that parameters, either in the form of an explicit "pro-drop parameter" in GB Theory or as manifested by the presence of pronominal features on T in Minimalism, are ineffective in accounting for language-internal microvariation. On the other hand, the OT practice of using violable constraints, and ranking them in different orders as appropriate to create variation, is capable of supporting explanations of intra-linguistic differences in null subject usage. The following passage summarises the crucial differences between constraints and parameters, and how the former are able to capture the essence of variability within a single language:

"Constraint violability is a very different thing from parameterization. A parameter describes a requirement that is either reliably enforced or completely ignored. [...] A constraint, no matter where it is ranked, always asserts its preference. [...] Whether it visibly asserts that preference depends on details of the language-particular ranking and the candidates under evaluation. [...] Even within a language, [...] a constraint might be active sometimes and inactive otherwise. This middle ground of partial activity follows from the interactional nature of OT, but it is difficult or impossible to achieve in parametric models."

(McCarthy 2002: 11-12)
Whereas parameters demand absoluteness by asserting without exception either the observation or ignoring of a certain linguistic condition, constraints allow for flexible variation by being inherently violable by nature, and by not being rigid cross-linguistically in their rankings relative to competing factors. Therefore, it follows rather convincingly that a linguistic phenomenon such as null subject usage, which displays both intra-linguistic as well as inter-linguistic variation, is best characterised not by a model involving parameters but instead by one consisting of violable constraints.

On an intuitive level, the rankings of the languages and of the features affecting pro-drop that we explored in Chapter 4 are most easily grasped within an OT framework. If we accept the reasonable hypothesis that certain Romance varieties in our study are more likely to license the null subject than others, and that within any one of those languages, there are certain morphosyntactic environments that are more conducive to tolerating covert subject pronouns than others, it is not irrational to argue that these differences all arise as a result of various constraint orderings. Those constraints that encourage phonetically null subjects are ranked higher in languages that show more null subject activity; it may be said as well that constraints impeding null subject usage are prioritised in the grammars of those languages that are less inclined to license pro. On a parallel note, when we look at the situation in a pro-friendly language, constraints relating to morphosyntactic conditions that are favourable to covert subject pronouns are promoted in the EVAL subcomponent of the architecture, while constraints linked to morphosyntactic conditions that are harsher to pro are relegated to lower positions. The opposite, of course, holds true for languages that are more resistant to pro; constraints tied to those morphosyntactic conditions that discourage the null subject are given prominent status in these linguistic systems.

Furthermore, the relative strengths of the numerous factors and features affecting null subjecthood are explicitly declared through their respective constraint rankings in each of the languages. If we see that, for instance, a constraint involving topichood is ranked higher in a grammar than a constraint detailing the presence of unique person agreement (as is likely to be the case for Puerto Rican Spanish, as we shall review later in this chapter), we can make a strong case that topichood has a greater influence on null subject tolerance than factors relating to person agreement, in that language. The clarity of such a conceptual representation, and the ease with
which we are able to associate higher constraint rankings with greater influence and relevance, are among the benefits of using OT to account for our data. Additionally, an OT model provides a straightforward method of investigating cross-linguistic tendencies, and perhaps even universal principles as well, by facilitating the comparison of the EVAL components of different languages and by subsequently allowing us to track consistent or unanimous trends in constraint rankings. Finally, we saw in the previous chapter that our attempts to reconcile our data with a GB framework resulted in a worrying escalation of parameters; this is echoed in the endorsement of constraint rankings below:

“Parameters are either on or off, but ranking provides much greater control over constraint activity. Similar control can be obtained in parametric theories only by proliferating parameters, eventually to absurdity”.

(McCarthy 2002: 110)

6.1.2 Making clear our stance

When we investigated the success of addressing our data within a Minimalist framework, we quickly encountered some of the same theoretical problems involving language-internal variation that had appeared in our investigation of the GB approach. Furthermore, our revision directed us towards the possibility of feature matrices being complex structures, a consideration presently given little attention and possessing inconclusive validity. While we are not in a position to argue that the geometry of feature bundles is indeed more intricate than previously thought (although our data do hint at its likelihood), our stance here is that much more research needs to be devoted to this area; before that occurs, it is risky, and perhaps even foolish, at this stage to construct a theory of null subjecthood on a linguistic premise whose foundation is currently insecure at best.

We hope to have justified the reasons for which we have opted to pursue an OT account of Romance null subject usage in the present work. It is worth offering here the following disclaimer that we are not ideologically biased in favour of using OT to explain syntax, or that we are unequivocally averse to Chomskyan syntactic
models. Indeed, we accept and agree that there are certain syntactic phenomena that cannot be adequately handled by OT, but rather are better addressed through a PPT approach, either in its GB or Minimalist manifestations. We have simply selected to account for null subject usage via an OT framework primarily on the basis of our results in Chapter 4. The items presented there did not conform neatly to either the GB or MP models, and the reworkings that would have been required to accommodate our data adequately were either unrealistic, unjustifiable, or impossible. It is merely at that point that we have concluded that applying OT to our data is a worthy endeavour. In sum, the move to appeal to OT in this work is not necessarily a blanket endorsement of it over PPT in every area of syntax; it is particular to the linguistic phenomena under investigation in the present work, and may not be applicable to other areas of syntactic study.

A more rigorous and specific defence of our decision to treat pro-drop in the context of OT instead of traditional Chomsky\textsuperscript{an} syntactic approaches is deferred until after we have introduced our original contribution to the field: a revised model of null subjecthood in an OT framework.

6.2 Pre-existing null subject explanations in OT

Before we offer our own version of an OT framework of null subject behaviour, it would be a beneficial exercise to review similar models that have been previously presented, which form the foundation upon which we base our approach. We anticipate that by indicating the drawbacks of these precursor models, we will be able to identify loci in need of improvement. This can only strengthen the validity of our version by drawing attention to those areas which have been neglected or ill-formulated, and which we attempt to mend in our version. Simultaneously, it justifies our motivation for being in a position to offer our own alternative proposal in the first place.
6.2.1 Deficiencies in Samek-Lodovici's (1996) model

We first review the model of null subjecthood as outlined in Samek-Lodovici (1996). The problems with his rendition of pro-drop are two-fold: not only does it contradict some of our results, but also it is not specific enough to account for differences witnessed in our data. Samek-Lodovici's model would have to be supplemented by additional constraints in order to explain adequately the microvariation present; his framework, as it currently stands, would not be able to capture this.

Samek-Lodovici proposes four constraints that affect null subjecthood: DROPTOPIC, PARSE, SUBJECT, and FULL-INT, all described in depth in Section 3.4.3. In languages displaying null subject activity, he asserts that the necessary ranking of these constraints is: DROPTOPIC >> PARSE >> SUBJECT, with FULL-INT ranked higher than SUBJECT (Samek-Lodovici 1996: 52; 57). So DROPTOPIC, which requires arguments having a topical antecedent to be left phonetically null (Samek-Lodovici 1996: 31), is the highest ranking of the four relevant constraints, meaning that its violation would be fatal to a candidate for output. However, we have data permitting lexicalised subject pronouns that are co-referenced with a topic antecedent:

(1a) [3it] La madre\textsubscript{1} di Gianni\textsubscript{2}, (pro\textsubscript{1}) lo\textsubscript{2} odi\textsubscript{a}.

(1b) La madre\textsubscript{1} di Gianni\textsubscript{2}, lei\textsubscript{1} lo\textsubscript{2} odi\textsubscript{a}.

As we have asserted before, the possibility that the overt subject pronoun may not be technically optional in this situation is not a grammatical concern; it is an issue for stylistics, pragmatics, and unmarkedness, the explanations for which are not our primary goals in the present work. What is relevant to note is that (1a), with an overt subject pronoun, is grammatically viable. However, (1b) violates DROPTOPIC because lei is co-referenced with madre, thus failing to observe the stipulation in DROPTOPIC that an argument whose antecedent is a topic is to be left unrealised. Therefore, Samek-Lodovici's framework predicts that (1b) should be ungrammatical, which is clearly not true.
We also potentially observe violations in the other direction. Whereas (1b) displayed an instance where a subject pronoun, co-referenced with a topic antecedent, was not left unrealised, we see in (2) below a situation where a subject pronoun, whose co-referenced antecedent is not a topic, has been left phonetically null:

(2) [22cs]  *Los vecinos₂ de Juan₁, sabemos que (pro₁) los₂ odia.

The antecedent of pro is co-referenced with the non-topic possessive antecedent Juan in the topicalised phrase. Whether this is strictly a violation of DROPTOPIC is not completely clear. If we go by Samek-Lodovici’s initial introduction and definition of the constraint, then (2) does not technically violate DROPTOPIC, because the constraint only asserts that arguments having topic antecedents be left unrealised; it makes no mention of those arguments that are co-referenced with non-topic antecedents. However, Samek-Lodovici offers a generalised corollary to the constraint, claiming that “null subjects must be licensed by topic antecedents” (1996: 37). If we are to interpret this to be a related or implicit component of the initial definition of the DROPTOPIC constraint, then we may reasonably argue that (2) does present a violation of this highest-ranked constraint. In this case, Samek-Lodovici’s framework predicts that (2) is grammatically unacceptable, contrary to what is actually the case.

The second principal drawback of Samek-Lodovici’s model is that it is unable to cope satisfactorily with some of the microvariation present in our data. Based solely on the four constraints deemed relevant to null subjecthood in his framework, we are at a loss to explain certain observable differences in the licensing of pro. The items below provide just one of several examples of this:

(3) [22ep]  *Os vizinhos₂ do João₁, sabemos que (pro₁) os₂ odeia.

(4a) [4ep]  *A mãe₂ do João₁, *pro₁ odeia-a₂.

(4b)  A mãe₂ do João₁, ele₁ odeia-a₂.
Going only by the four constraints that Samek-Lodovici offers, we expect that (3) and (4a) should show identical null subject activity because their computations according to EVAL are exactly alike. However, they do not; as (4b) shows, pro is not an option for that sentence, whereas it is for (3). There is no explanation as to why this should be so. Both (3) and (4a) violate DROPTOPIC, the highest-ranked constraint. Violations occur as well with PARSE by not lexicalising the subject pronoun, and with SUBJECT by leaving the highest A-position of the clause unrealised. FULL-INT, which applies only to expletives, is equally irrelevant. Despite identical evaluations by the four constraints present, we see opposite licensings of pro. In this particular situation, we point out that the difference between (3) and (4) is the intervening main clause in the former, and we claim that it does have a tangible effect on the tolerance of pro. This factor is not incorporated into SamekLodovici’s system of four constraints, hence the grammar’s inability to discriminate accurately between the two sentences. What this exposition above signifies is that one of three solutions (or some combination thereof) is needed to eliminate these apparent discrepancies. Either we need to reformulate the definition of DROPTOPIC, reconsider the rankings of the constraints relevant to null subject licensing, or add new constraints to cover data and factors not considered in Samek-Lodovici’s work.

We now examine to what extent these grievances are sufficiently addressed in Cole’s (2000) proposal, which he offers as a revised version of Samek-Lodovici’s model.

### 6.2.2 Deficiencies in Cole’s (2000) model

The principal difference between Cole’s and Samek-Lodovici’s models is that Cole incorporates into his constraints semantic adequacy, antecedent competition, and morphological maximality, issues not covered in Samek-Lodovici’s work. We claim that these factors do not need to be included in the definitions of Cole’s constraints, but instead are discernible from the innovative constraints to be proposed in our work, and from how these interact with each other. Furthermore, some of his constraints either require reworking of their definitions, or may be dispensed with altogether as being obsolete or irrelevant.

The fundamental constraint in Cole’s framework, upon which he models his
other constraints, is offered as an improvement over Samek-Lodovici’s DROPTOPIC. Labelled DROPSUBJECT, it states, “do not realise unfocused pronominal subject co-refering with the last recent sentence topic that provides a semantically adequate subject to the verb of the current sentence” (Cole 2000: 226). Cole’s DROPSUBJECT qua revision of Samek-Lodovici’s DROPTOPIC takes into account Cole’s work on antecedent competition and syntactic structures that Samek-Lodovici does not examine. The part of DROPSUBJECT’s definition mentioning “last recent sentence topic” merely reflects Cole’s investigation of the two-sentence sequence that is one of his own original contributions to the field of research, and that we adopt with modifications in the present research. The inclusion of the term “semantically adequate” in DROPSUBJECT, not present in Samek-Lodovici’s DROPTOPIC, is problematic, as we shall expand upon below.

Despite Cole’s amendments, we argue that DROPSUBJECT is just a cosmetic transformation of DROPTOPIC; upon deeper inspection, we notice that DROPSUBJECT is not in fact a significant improvement over Samek-Lodovici’s DROPTOPIC. This is due to the fact that the problems affecting DROPTOPIC are also applicable to DROPSUBJECT. Just as we saw in Section 6.2.1 above with DROPTOPIC, there are instances in our data of unfocused subjects that do co-refer with the last recent sentence topic but that are in fact structurally realised, thus violating DROPSUBJECT in parallel fashion. The marking of sentences containing overt subject pronouns co-referent with previous topics as ungrammatical is another fault that is common to both Cole’s and Samek-Lodovici’s models, so in this respect DROPSUBJECT is not clearly preferable to DROPTOPIC.

The definition of DROPSUBJECT contains the term “semantically adequate”, in reference to possible antecedents that can serve as a co-referent to the subject of the verb under consideration. This is another inclusion that Cole offers as an amelioration to Samek-Lodovici’s DROPTOPIC. However, semantically adequacy is vague and too open to interpretation, and needs to be better defined. Therefore, one of our aims is to attempt a formal description of semantic adequacy; we hope to achieve this by deconstructing this concept along three lines of identification types: person, number, and gender features. Our goal is to make explicit how exactly antecedents qualify as being semantically adequate.

One apparent benefit of DROPSUBJECT is that Cole does not make any mention of co-reference with non-topic antecedents, an unambiguity not clarified in
Samek-Lodovici's version of this constraint. However, we still run into problems because DROPSUBJECT is unable to discern between (5) and (6) below:

(5) [32bp] \((\text{pro}_1) \text{Encontro-me com o meu pai}_2 \text{ hoje.} (\text{pro}_2) \text{Odiava-me}_1.\)

(6) [28bp] \(\text{O João}_1 \text{ se encontra com a Maria}_2 \text{ hoje.} *\text{pro}_2 \text{Odeia ele}_1.\)

Both sentences are identical with respect to DROPSUBJECT, PARSE, and SUBJECT, yet they differ in their grammaticality. Cole's framework is unable to account for this divergence.

So it is difficult to detect tangible points where DROPSUBJECT really is an improvement over DROPTOPIC. We will need to define this constraint in a theoretically rigorous manner, such that it will not suffer from the same shortcomings that inhibit the effectiveness of DROPSUBJECT and DROPTOPIC.

We now address some of Cole's other constraints, whose formulations are rooted in his original DROPSUBJECT. DROPSUBJECT3 involves morphological maximality, Cole's attempt at formalising the link between features identifiable by morphology and the extent to which pro is licensed. In the next section we will attempt to show that, in our proposed framework, morphological maximality does not need to be mentioned within the definition of a constraint. Indeed, we question to what extent it is even relevant to null subject licensing.

DROPSUBJECT5 and DROPSUBJECT6 are directed towards the issue of antecedent competition in two separate syntactic constructions under investigation by Cole. The first constraint applies to the two-sentence [2S] sequences, the second to the [T, MC] and [T, MC, SC] structures; both constraints deal with data containing subjects that are necessarily co-referenced with the possessor (i.e., the non-topic antecedent) because it is the only possible antecedent "by default". That is to say, the previous topic is both syntactically and logically unavailable, due to its co-referencing with a direct object pronoun, to serve as the antecedent for the subject under inspection. As an alternative to specifying "co-referencing by default" within the definition of the constraints, we choose to formalise this phenomenon through the constraints themselves and their interactions. We find it more beneficial to the theoretical sustainability of our approach to eliminate explicit mention of default,
instead opting to model this through the mechanism available to us in EVAL. We hypothesise that default co-referencing will emerge elegantly from the observation or violation of our new constraints that address the number of competing antecedents present based on feature-indicating morphology.

As for the remainder of the constraints that constitute Cole's framework, we shall clarify in the next section our decision to discard the Baković-motivated FAITH[SUB], a move based on essential differences between Cole's methodological approach and ours. We have already discussed in Section 3.7.5 our reasons for not considering DROPSUBJECT2 and DROPSUBJECT4 in our analysis.

With a clearer awareness of the faults within Samek-Lodovici's and Cole's proposals, and of the need to improve upon their models, we understand better now which problems must be resolved in order to create a tenable framework, and how to go about accomplishing this. Keeping in mind our goal of satisfactorily addressing these shortcomings while successfully accounting for all of our data, we now present our revised model of null subjecthood in an Optimality Theoretic framework.

6.3 A new proposal

Our principal original contribution to the study of null subject usage is developed here. We begin by describing the constraints that will feature in our model, and then move on to an examination of how the interactions of these constraints are able to account for our data. We then describe all of the Romance varieties under consideration in our work by assigning each of them a unique system of constraint rankings that appropriately reflects their specific null subject properties.

6.3.1 Defining the constraints

6.3.1.1 DROPGENDER, DROPNUMBER, DROPPERSON, and DROPNODIFF

We witnessed in Section 6.2 that Samek-Lodovici's and Cole's constraints, DROPTOPIC and DROPSUBJECT respectively, were inadequate in a number of ways. Not only did they lead occasionally to incorrect predictions of
instead opting to model this through the mechanism available to us in EVAL. We hypothesise that default co-referencing will emerge elegantly from the observation or violation of our new constraints that address the number of competing antecedents present based on feature-indicating morphology.

As for the remainder of the constraints that constitute Cole’s framework, we shall clarify in the next section our decision to discard the Baković-motivated FAITH[SUB], a move based on essential differences between Cole’s methodological approach and ours. We have already discussed in Section 3.7.5 our reasons for not considering DROPSUBJECT2 and DROPSUBJECT4 in our analysis.

With a clearer awareness of the faults within Samek-Lodovici’s and Cole’s proposals, and of the need to improve upon their models, we understand better now which problems must be resolved in order to create a tenable framework, and how to go about accomplishing this. Keeping in mind our goal of satisfactorily addressing these shortcomings while successfully accounting for all of our data, we now present our revised model of null subjecthood in an Optimality Theoretic framework.

6.3 A new proposal

Our principal original contribution to the study of null subject usage is developed here. We begin by describing the constraints that will feature in our model, and then move on to an examination of how the interactions of these constraints are able to account for our data. We then describe all of the Romance varieties under consideration in our work by assigning each of them a unique system of constraint rankings that appropriately reflects their specific null subject properties.

6.3.1 Defining the constraints

6.3.1.1 DROP(GENDER), DROP(NUMBER), DROP(PERSON), and DROP(NO_DIFF)

We witnessed in Section 6.2 that Samek-Lodovici’s and Cole’s constraints, DROPTOPIC and DROPSUBJECT respectively, were inadequate in a number of ways. Not only did they lead occasionally to incorrect predictions of
ungrammaticality, but also they were not specific enough to account for all of the patterns observed in our data.

The crucial shortcoming of DROPTOPIC and DROPSUBJECT can be found in their focus on whether the subject is co-referenced with a topic or non-topic antecedent. We refer back to the odd-numbered sentences presented in Chapter 4, and note two things: first, all of them are instances of co-reference with a topic antecedent, and second, all of them are grammatical, regardless of whether the subject is overt or covert. In other words, in all of the sentences where co-reference occurs between the subject and a topic antecedent, both the version of the sentence with pro, and the one without it, are grammatical possibilities. Therefore, any constraint that bases its principal schism on whether or not the co-referenced antecedent is a topic will not be informative to us, since our data show that pro will always be tolerated if it is co-referenced with the topic antecedent. Where the microvariation lies, therefore, is within the subset of those sentences containing a subject co-referenced with the non-topic antecedent, and so we need constraints that will look at specific distinctions occurring within the realm of switch reference.

It can be discerned from the data in Chapter 4 that the predominant factor affecting null subject usage is the presence of antecedent-disambiguating morphology. This concept can be decomposed further, since we are aware that the two antecedents can be discriminated from each other by at least one of three different features, and that these three features do not all act uniformly. Therefore, we may assert specifically that gender-, number-, and person-distinguishing morphologies are the essential determiners of null subjecthood.

Our proposed replacements to DROPTOPIC and DROPSUBJECT embody this approach. We thus choose to incorporate into the system factors relating to antecedent-disambiguating morphology instead of topichood. Our new constraints are as follows:
(7) DROP(GENDER): Drop pronoun when antecedent of subject is uniquely identified by gender-indicating morphology

Observed: [28cs]  
Juan, va a quedar con María hoy. (pro₂) Le₁ odiₐ.

Violated: [28pr]  
Juan, va a encontrarse con María hoy. Ella₂ lo₁ odiₐ.

(8) DROP(NUMBER): Drop pronoun when antecedent of subject is uniquely identified by number-indicating morphology

Observed: [30ep]  
O João₁ encontra-se com os vizinhos₂ hoje. (pro₂) Odeiam-no₁.

Violated: [30bp]  
O João₁ se encontra com os vizinhos₂ hoje. Eles₂ odeiam ele₁.

(9) DROP(PERSON): Drop pronoun when antecedent of subject is uniquely identified by number-indicating morphology

Observed: [32it]  
(pro₁) Incontro mio padre₂ oggi. (pro₂) Mi₁ odiava.

Violated: [32pr]  
(pro₁) Voy a encontrarme con mi padre₂ hoy. El₂ me₁ odiaba.

To cover all situations, we must also create a fourth constraint:

(10) DROP(NO_DIFF): Drop pronoun when the antecedents are indistinguishable by any feature

Observed: [26ct]  
En Joan₁ es reuneix amb en Miquel₂ avui. (pro₂) el₁ menyspreua.

Violated: [26cs]  
Juan₁ va a quedar con Miguel₂ hoy. Él₂ le₁ odiₐ.

DROP(NO_DIFF) accounts for those instances where the other three DROP constraints are fulfilled vacuously. If the two antecedents mirror each other in gender, number, and person, then none of the feature-specific DROP constraints can technically be violated, since they are not applicable in the evaluation. However, we need to be able to explain the difference between Catalan example (26ct), where pro is tolerated, and its Castilian counterpart (26cs), where an overt subject pronoun is required; DROP(NO_DIFF) is able to accomplish this.

---

1 We repeat here that these indications can occur anywhere in the predicate, either directly (e.g., subject-verb agreement makes possible only one of the two antecedents) or indirectly (e.g., object pronoun is necessarily co-referenced with one of the antecedents, leaving the other the only remaining antecedent for co-reference with the subject).
(7) DROP(GENDER): Drop pronoun when antecedent of subject is uniquely identified by gender-indicating morphology

Observed: [28cs]  Juan₁ va a quedar con María₂ hoy. (pro₂) Le₁ odia.
Violated: [28pr]  Juan₁ va a encontrarse con María₂ hoy. Ella₂ lo₁ odia.

(8) DROP(NUMBER): Drop pronoun when antecedent of subject is uniquely identified by number-indicating morphology

Observed: [30ep]  O João₁ encontra-se com os vizinhos₂ hoje.
                (pro₂) Odeiam-no₁.
Violated: [30bp]  O João₁ se encontra com os vizinhos₂ hoje. Eles₂ odeiam ele₁.

(9) DROP(PERSON): Drop pronoun when antecedent of subject is uniquely identified by number-indicating morphology

Observed: [32it]  (pro₁) Incontro mio padre₂ oggi. (pro₂) Mi₁ odiava.
Violated: [32pr]  (pro₁) Voy a encontrarme con mi padre₂ hoy. El₂ me₁ odiaba.

To cover all situations, we must also create a fourth constraint:

(10) DROP(NO_DIFF): Drop pronoun when the antecedents are indistinguishable by any feature

Observed: [26ct]  En Joan₁ es reuneix amb en Miquel₂ avui. (pro₂) el₁ menyspreua.
Violated: [26cs]  Juan₁ va a quedar con Miguel₂ hoy. Él₂ le₁ odia.

DROP(NO_DIFF) accounts for those instances where the other three DROP constraints are fulfilled vacuously. If the two antecedents mirror each other in gender, number, and person, then none of the feature-specific DROP constraints can technically be violated, since they are not applicable in the evaluation. However, we need to be able to explain the difference between Catalan example (26ct), where pro is tolerated, and its Castilian counterpart (26cs), where an overt subject pronoun is required; DROP(NO_DIFF) is able to accomplish this.

---

¹ We repeat here that these indications can occur anywhere in the predicate, either directly (e.g., subject-verb agreement makes possible only one of the two antecedents) or indirectly (e.g., object pronoun is necessarily co-referenced with one of the antecedents, leaving the other the only remaining antecedent for co-reference with the subject).
6.3.1.2 PARSE(PRO), PARSE(C), and PARSE(S)

Both Samek-Lodovici and Cole include in their frameworks a PARSE constraint, which is observed whenever all input items are structurally realised into the phrase structure. PARSE is violated whenever an item, such as a subject pronoun, is left unrealised. Thus, PARSE incurs a violation whenever pro survives into the output at the expense of an overt subject pronoun, and is the critical constraint that determines whether the null subject can be licensed. However, we cover in our work syntactic structures not investigated by Samek-Lodovici, and Cole adopts Samek-Lodovici’s version of PARSE straight into his own analysis. Therefore, there is the need to expand on their formulations of PARSE to make it relevant to our framework.

We examined in our research three syntactic structures that differed according to the distance between the subject and its antecedent. Furthermore, it was observed that not all three syntactic structures under consideration licensed pro in the same manner. Therefore, it seems rational to include in our model a series of constraints that incorporate null subject tolerance as a function of the distance separating the subject and its antecedent. We thus postulate the following constraints:

(11) PARSE(PRO): Realise pronoun when subject and non-topic antecedent are in the same clause

Observed: [4it]  
La madre2 di Gianni1, lui1 la2 odia.

Violated: [4ct]  
*La mare2 d’en Joan1, pro1 la2 menyspreua.

(12) PARSE(C): Realise pronoun when subject and non-topic antecedent are separated by a clausal boundary

Observed: [22bp]  
Os vizinhos2 do João1, sabemos que ele1 odeia eles2.

Violated: [22ep]  
Os vizinhos2 do João1, sabemos que (pro1) os2 odeia.

(13) PARSE(S): Realise pronoun when subject and non-topic antecedent are separated by a sentential boundary

Observed: [28pr]  
Juan1 va a encontrarse con María2 hoy. Ella2 lo1 odia.

Violated: [28cs]  
Juan1 va a quedar con María2 hoy. (pro2) Le1 odia.
PARSE(PRO), PARSE(C), and PARSE(S) correspond to the structures that we have been thus far labelling as [T, MC], [T, MC, SC], and [2S] respectively.

We have included mention of non-topic antecedents in these constraint definitions to verbalise our position of investigating variation within switch-reference environments only, recalling from above that co-reference with a topic antecedent will never force the realisation of an overt subject pronoun.

For the purposes of our work, there is no way in which all three constraints may simultaneously be satisfied vacuously, since we only examine sentences containing one of the three syntactic structures represented by our PARSE constraints. Therefore, it will always be the case that one, and only one, constraint will be applicable in any evaluation, and so no default constraint comparable to DROP(NO_DIFF) for the DROP constraints need be postulated for the PARSE constraints.

6.3.1.3 A disclaimer on grammaticality, optionality, and Optimality Theory

Before we detail how our seven constraints\(^2\) interact to account for the intra- and inter-linguistic microvariation in null subject usage present in our data, it would be worth clarifying here our stance on suboptimal candidates in OT and their grammaticality.

There is discrepancy in the OT literature on whether the optimal candidate selected by EVAL is the best (or most appropriate) structure, or whether it is the only grammatical one. These two results are not identical. Identifying the best or most appropriate structure suggests the possibility that there may be at least one losing candidate that is still grammatical, albeit not optimal in light of factors relating to the nature of, or conditions on, the input. Asserting that the output given by EVAL is the only grammatical one necessarily implies that all suboptimal candidates are

\(^2\) We note that Samek-Lodovici and Cole use two constraints that we have chosen not to incorporate into our model. SUBJECT deals with whether the subject pronoun appears preverbally or postverbally. While we do not deny that postponing the subject after the verbal unit can have relevant, non-trivial ramifications on the licensing of pro, we have not investigated postverbal versions of our data items. FULL-INT, present only in Samek-Lodovici’s framework, involves expletives. We maintain that postverbal and expletive subjects are outside the immediate purview of our research, so we do not include these two constraints in our approach.
ungrammatical.

Samek-Lodovici follows Prince and Smolensky (1993) in declaring that “all candidates which are suboptimal are ungrammatical” (1996: 3). Cole writes that “the optimal candidate is the only grammatical one” (2000: 205). However, McCarthy states that “the actual output – the most harmonic or optimal candidate – is the one that is more harmonic in all its pairwise competitions with other candidates” (2002: 3). We note his position that suboptimal structures are less harmonic than the winning candidate; it is not claimed that they are strictly ungrammatical.

It is acknowledged that we have been critical of Samek-Lodovici’s and Cole’s models because grammatical candidates are occasionally eliminated as suboptimal by EVAL. Our main complaint, however, is not so much with this apparent defect in Optimality Theory itself, but rather with Samek-Lodovici’s and Cole’s stance on this issue. By claiming that only the optimal candidate is grammatical, they have committed to a viewpoint whereby all losing structures are ungrammatical, which cannot be tenable. If this were the case, then all work done in Optimality Theory on optionality, such as those on that-deletion by Baković (1997), explored in Section 3.6, and on wh-phrase deletion by Baković and Keer (2001), would be fundamentally flawed because these cases all deal with choosing between two perfectly grammatical options, based solely on constraints ruling on the faithfulness of the particular input. McCarthy (2002: 201; 230) concedes that this controversy is ubiquitous, and that a clean, satisfactory resolution of this may be unattainable due to the inherent nature of the OT mechanism.

We therefore have to adopt in the present research the liberal interpretation that suboptimal candidates are not necessarily ungrammatical, just dispreferred in some respect. It is admitted that some may find this to be a significant weakening in our model, but our decision is empirically supported by the data. To do otherwise would both destroy the foundations of OT explanations of optionality and also fail to relieve us of some of the problems in Samek-Lodovici’s and Cole’s models. This may be an unfortunate side effect of choosing to treat null subjecthood in an OT framework, but it is perhaps unavoidable.

An alternate option to treating this optionality conundrum would be to incorporate Baković’s method of introducing a FAITHFULNESS constraint into our model, so that in instances of co-reference with the topic antecedent, both the structure including pro and that containing the overt subject pronoun may be
predicted as grammatical depending on whether the lexical subject is present in the input. This poses two problems. First, FAITHFULNESS would need to rank above all of our seven constraints only in instances of co-reference with the topic antecedent, but it would need to rank at the bottom of the hierarchy in cases of co-reference with the non-topic antecedent in order to be dominated and rendered inactive. This is clearly anathematic to our notion of fixed constraints within EVAL to determine a language’s unique grammar. Second, it is not obvious that inserting pro into the input is the same as not inserting an overt subject pronoun, and whether either of these two options is incompatible with the mechanisms of Minimalist computation. For example, if an overt subject pronoun is absent in the input, but our OT grammar specifies its phonetic actualisation in the output, would its sudden appearance after Spell-Out violate Chomsky’s Inclusiveness condition (1995: 228)? Indeed, there does not seem to be any easy or clear solution to this optionality puzzle.

We must also yield to the possibility that, where grammatical or syntactic optionality exists in using pro, there may be pragmatic or discourse-related factors that can distinguish between the two competing structures. In this sense, of course, we are not witnessing true optionality then. However, as we are not considering pragmatics or discourse analysis in our framework, for the purposes of our specific inquiry, we must treat the two competing structures to be syntactically optional.

Our decision here will become pertinent when we delve into the operations and details of our proposed OT model. In particular, there will be instances when the grammar selects a sentence with pro over its counterpart with an overt subject pronoun, even though both are permissible. It will soon become clear that this arises as a result of our approach to the problem. Whereas Samek-Lodovici looked at which structure was stylistically optimal, and Cole examined where one had the option of using pro, we have investigated where one must use an overt subject pronoun because ungrammaticality results if pro is present. Therefore, the focus of our model will be on determining when pro cannot be used, and not so much on how to rectify this optionality/dual grammaticality debate.
6.3.2 Interaction of constraints

We now examine how the interaction of our seven constraints produces various outcomes. It is best to consider this from the perspective of interactions between constraints from two different sets, the DROP set and the PARSE set. Whenever a certain DROP constraint is ranked higher than a particular PARSE constraint, this indicates that the null subject is licensed in the morphosyntactic environment specified by the constraints. If PARSE outranks DROP, then pro is not available; an overt subject pronoun must be used. This interaction between DROP and PARSE is consistent with what is observed with DROPTOPIC/DROPSUBJECT and PARSE in Samek-Lodovici’s and Cole’s frameworks.

In our model, when DROP(x) outranks PARSE(y), that signifies that when the two antecedents can be disambiguated through x-indicating morphology in the syntactic environment ‘y’, the null subject is permitted. The alternate ranking means that the null subject is prohibited, and that an overt subject pronoun is required.

It should be mentioned that, because we are primarily concerned only with whether the overt subject must be used, there are only two candidates in each evaluation, one with pro, the other with the lexical subject. Furthermore, we must add the stipulation that whenever a structure with pro is selected as the optimal candidate, its counterpart with an overt subject pronoun is not ungrammatical, only that it is less harmonious with the constraints than the winning candidate. The reverse does not hold true; when a structure with an overt subject pronoun is selected as optimal, this implies that the equivalent with pro is ungrammatical. The default, then, is that the option to use a null subject exists unless our OT model selects a structure with an overt subject pronoun as the winning candidate. This also reflects the empirical fact that null subjects are the more frequent, unmarked option. So when our model identifies a null subject structure as the optimal candidate, we have the option of using an overt subject pronoun; when our model selects a structure with an overt subject pronoun, its presence is a grammatical requirement. This qualification may hinder the meticulousness of our framework, but we agree with McCarthy that this situation seems inescapable, and claim that the benefits of treating null subjecthood in OT instead of in PPT outweigh whatever distractions arise as a result of this.

Returning to the mechanism of our model, we would need to examine pair-wise interactions of the four constraints from the DROP group and the three from the
PARSE set, keeping in mind that when DROP outranks PARSE, pro is licensed, and when PARSE outranks DROP, an overt subject pronoun is required. The twelve binary interactions are therefore:

(a) DROP(NO_DIFF)/PARSE(PRO)
(b) DROP(GENDER)/PARSE(PRO)
(c) DROP(NUMBER)/PARSE(PRO)
(d) DROP(PERSON)/PARSE(PRO)
(e) DROP(NO_DIFF)/PARSE(C)
(f) DROP(GENDER)/PARSE(C)
(g) DROP(NUMBER)/PARSE(C)
(h) DROP(PERSON)/PARSE(C)
(i) DROP(NO_DIFF)/PARSE(S)
(j) DROP(GENDER)/PARSE(S)
(k) DROP(NUMBER)/PARSE(S)
(l) DROP(PERSON)/PARSE(S)

There is no benefit gained from comparing the interactions among the DROP constraints, or the same among the PARSE constraints, because members of these two groups must necessarily interact with members of the other group to be informative. The only comparisons we will make within the PARSE group or within the DROP group of constraints will take place in the final rankings of all seven constraints for each language, to determine the language-specific effects that different syntactic environments, and different morphological distinctions, have on the licensing of pro.

We shall now briefly, but systematically, review below all twelve of these interactions, and how our data from Chapter 4 fit into them.

(a) DROP(NO_DIFF)/PARSE(PRO): The two antecedents are not distinguishable by any features, and are in the same clause as their potential co-referent. This corresponds to structures (2) in Section 4.1.1. All six versions of (2) require an overt

---

3 This constraint pair also corresponds to structures (1) in that same section, which are the corresponding versions of (2) but with reference to the topic antecedent. We choose to deal with these structures later, for two reasons. First, the present model is geared towards switch reference instances, so it seems best to develop our model initially with respect to those structures. Once that has been accomplished, it will become clear how to make the model compatible with instances of co-reference with the topic antecedent; that is the opportune moment to address the odd-numbered structures. Therefore, throughout section 6.3.2, we shall only make specific reference to the even-numbered structures, where co-reference is with the non-topic antecedent.
subject pronoun, so for all of our languages, PARSE (PRO) outranks DROP(NO_DIFF).

DROP(NO_DIFF) >> PARSE(PRO): none
PARSE(PRO) >> DROP(NO_DIFF): BP, CS, CT, EP, IT, PR

(b) DROP(GENDER)/PARSE(PRO): The two antecedents are distinguishable by gender only, and are in the same clause as their potential co-referent. This corresponds to structures (4) in Section 4.1.2. All six versions of (4) require an overt subject pronoun, so for all of our languages, PARSE (PRO) outranks DROP(GENDER).

DROP(GENDER) >> PARSE(PRO): none
PARSE(PRO) >> DROP(GENDER): BP, CS, CT, EP, IT, PR

(c) DROP(NUMBER)/PARSE(PRO): The two antecedents are distinguishable by number only, and are in the same clause as their potential co-referent. This corresponds to structures (6) in Section 4.1.3. All six versions of (6) require an overt subject pronoun, so for all of our languages, PARSE (PRO) outranks DROP(NUMBER).

DROP(NUMBER) >> PARSE(PRO): none
PARSE(PRO) >> DROP(NUMBER): BP, CS, CT, EP, IT, PR

(d) DROP(PERSON)/PARSE(PRO): The two antecedents are distinguishable by person only, and are in the same clause as their potential co-referent. This corresponds to structures (8) in Section 4.1.4. Five of the six versions of (8) permit pro and thus rank DROP(PERSON) above PARSE(PRO). Puerto Rican Spanish requires an overt subject pronoun, so PARSE (PRO) outranks DROP(PERSON).

DROP(PERSON) >> PARSE(PRO): BP, CS, CT, EP, IT
PARSE(PRO) >> DROP(PERSON): PR

We now repeat the procedure, substituting PARSE(C) for PARSE(PRO). The syntactic environment under consideration now involves an intervening clause between the subject and its antecedents.

(e) DROP(NO_DIFF)/PARSE(C): The two antecedents are not distinguishable by any features, and are separated from their potential co-referent by a clausal boundary. This corresponds to structures (18) in Section 4.2.1. All six versions of (18) require an overt subject pronoun, so for all of our languages, PARSE(C) outranks DROP(NO_DIFF).
The reader should be comfortable now with the procedure for the analysis of the seven remaining constraint pairs. In the interest of space and efficiency, we streamline them below, indicating only the corresponding structures from Chapter 4 and which languages order the constraints in which direction.

(f) DROP(GENDER)/PARSE(C)
Sentences (20) from Section 4.2.2
DROP(GENDER) >> PARSE(C): none
PARSE(C) >> DROP(GENDER): BP, CS, CT, EP, IT, PR

(g) DROP(NUMBER)/PARSE(C)
Sentences (22) from Section 4.2.3
DROP(NUMBER) >> PARSE(C): CS, CT, EP
PARSE(C) >> DROP(NUMBER): BP, IT, PR

(h) DROP(PERSON)/PARSE(C)
Sentences (24) from Section 4.2.4
DROP(PERSON) >> PARSE(C): BP, CS, CT, EP, IT
PARSE(C) >> DROP(NUMBER): PR

Analyses of the final four constraint pairs proceed in similar fashion, the only difference being that PARSE(S) has now replaced PARSE(C), so there is a sentential boundary between the competing antecedents and the potential co-referenced subject.

(i) DROP(NO_DIFF)/PARSE(S)
Sentences (26) from Section 4.3.1
DROP(NO_DIFF) >> PARSE(S): CT
PARSE(S) >> DROP(NO_DIFF): BP, CS, EP, IT, PR

(j) DROP(GENDER)/PARSE(S)
Sentences (28) from Section 4.3.2
DROP(GENDER) >> PARSE(S): CS, CT
PARSE(S) >> DROP(GENDER): BP, EP, IT, PR

(k) DROP(NUMBER)/PARSE(S)
Sentences (30) from Section 4.3.3
DROP(NUMBER) >> PARSE(S): CS, CT, EP, IT
PARSE(S) >> DROP(NUMBER): BP, PR
Most of our data from Chapter 4\(^4\) have been accounted for here by means of the interactions of our original constraints. We now try to make sense of them in the context of each language.

### 6.3.3 Using the constraints to understand the data

We have categorised above for each of the twelve pair-wise constraint interactions which languages rank the PARSE constraint higher, and which rank the DROP constraint higher. However, it may be more insightful instead to sort these constraint interactions by language. This will assist us in attaining a better appreciation of the ramifications of these constraint rankings on each individual language. It is important to keep in mind that we have been limiting instances of coreference in Section 6.3 thus far to those with the non-topic antecedent, so the analyses performed below should be interpreted solely in this restricted context for the present moment.

#### 6.3.3.1 Brazilian Portuguese

We list below the twelve constraint rankings outlined in Section 6.3.2 that pertain to our Brazilian Portuguese data:

(a) PARSE(PRO) >> DROP(NO DIFF)
(b) PARSE(PRO) >> DROP(GENDER)
(c) PARSE(PRO) >> DROP(NUMBER)
(d) DROP(PERSON) >> PARSE(PRO)

---

\(^4\) Only the even-numbered structures from Sections 4.1 through 4.3, containing co-reference with the topic antecedent, and data from Section 4.4 on gender-inherent predicates, have yet to appear. They will do so in Section 6.4.5.
Before we list possible hierarchies of all seven constraints within the Brazilian Portuguese EVAL, let us pause for a moment to assess the meaning of these pair-wise constraint interactions. The first group of constraints, involving PARSE(PRO) and the DROP set, shows that PARSE(PRO) outranks all DROP constraints except for DROP(PERSON). This signifies that for the relevant syntactic environment (subject and antecedents in the same clause), a null subject will only be permitted if the two antecedents are distinguishable by person features. The opposite is true if the two competing antecedents can be discriminated by only gender or number, or are not differentiable at all; an overt subject pronoun is required.

The second and third groups of four can be considered similarly to the first group, the only changes being that there are different syntactic environments under consideration. In the second group with PARSE(C), there is a clausal boundary that separates the subject from its potential antecedents, while in the third group with PARSE(S), there is a sentential boundary present.

Again, we see that PARSE(C) and PARSE(S) outrank all DROP constraints except for DROP(PERSON), so the analysis used for PARSE(PRO) above is directly applicable to PARSE(C) and PARSE(S). This parallelism is nothing more than a coincidence in the Brazilian Portuguese grammar; we will not necessarily see this symmetry when we examine the grammars of our other Romance varieties. Our interpretation of the twelve pair-wise constraints could have been conducted in the other direction as well, by examining the interaction between a particular DROP constraint and the PARSE set, with the same final result.

We now attempt to formulate a ranking hierarchy of these seven constraints. Keeping in mind our observation that all the PARSE constraints outrank all the DROP constraints except for DROP(PERSON), here are some possible renderings of
Brazilian Portuguese EVAL:

DROP(PERSON) >> PARSE(PRO) >> PARSE(C) >> PARSE(S) >> DROP(NUMBER) >> DROP(GENDER) >> DROP(NO_DIFF)

DROP(PERSON) >> PARSE(S) >> PARSE(PRO) >> PARSE(C) >> DROP(GENDER) >> DROP(NO_DIFF) >> DROP(NUMBER)

DROP(PERSON) >> PARSE(C) >> PARSE(S) >> PARSE(PRO) >> DROP(NO_DIFF) >> DROP(NUMBER) >> DROP(GENDER)

Although we have listed above a number of potential ranking options for the grammar, it must be pointed out that Brazilian Portuguese EVAL is unique. In other words, it is not the case that all three of these are possible; in fact, only one can be. These are the most definite ranking orders that we can establish based on the limited empirical results collected in our work. Further research and data will eventually be able to eliminate all but one of the grammars that we have claimed are possible.

In light of the multiple potential grammars created by the dearth of our data, we are perhaps best able to summarise the ranking of constraints in Brazilian Portuguese EVAL as follows:

DROP(PERSON) >> \{PARSE(PRO), PARSE(C), PARSE(S)\} >> \{DROP(NUMBER), DROP(GENDER), DROP(NO_DIFF)\}

This representation is both consistent with all twelve pair-wise constraint interactions examined in this section, as well as accurate in capturing all possible grammars.

6.3.3.2 Castilian

The analyses for the remaining five Romance varieties under consideration in this work follow that outlined for Brazilian Portuguese above. Therefore, we
streamline their presentations below by presenting their twelve constraint interactions and a corresponding representation of their EVAL, making occasional comments as required or notable.

The twelve pair-wise constraint rankings for Castilian are as follows:

(a) PARSE(PRO) >> DROP(NO_DIFF)
(b) PARSE(PRO) >> DROP(GENDER)
(c) PARSE(PRO) >> DROP(NUMBER)
(d) DROP(PERSON) >> PARSE(PRO)
(e) PARSE(C) >> DROP(NO_DIFF)
(f) PARSE(C) >> DROP(GENDER)
(g) DROP(NUMBER) >> PARSE(C)
(h) DROP(PERSON) >> PARSE(C)
(i) PARSE(S) >> DROP(NO_DIFF)
(j) DROP(GENDER) >> PARSE(S)
(k) DROP(NUMBER) >> PARSE(S)
(l) DROP(PERSON) >> PARSE(S)

We notice here that the three PARSE constraints do not interact in the same manner with the DROP group; this distinction was neutralised in the Brazilian Portuguese analysis above. This is evidence that the three syntactic structures that we have been considering are indeed non-trivially distinct from each other; that they display different results lends credence to the proposition that distance between subject and antecedent does have a detectable effect on the licensing of pro. This will be explored in greater depth later.

The ranking of constraints is as follows:

DROP(PERSON) >> PARSE(PRO) >> DROP(NUMBER) >> PARSE(C) >> DROP(GENDER) >> PARSE(S) >> DROP(NO_DIFF)

The twelve constraint interactions for Castilian create just this one unique final ranking. No other orderings are possible. This gives us a clearer picture as well of the varying effects of the different features. We observe with the Castilian data that gender-, number-, and person-indicating features all display distinct strengths in
licensing *pro*.

### 6.3.3.3 Catalan

The twelve pair-wise constraint interactions for Catalan are as follows:

(a) PARSE(PRO) >> DROP(NO_DIFF)
(b) PARSE(PRO) >> DROP(GENDER)
(c) PARSE(PRO) >> DROP(NUMBER)
(d) DROP(PERSON) >> PARSE(PRO)

(e) PARSE(C) >> DROP(NO_DIFF)
(f) PARSE(C) >> DROP(GENDER)
(g) DROP(NUMBER) >> PARSE(C)
(h) DROP(PERSON) >> PARSE(C)

(i) DROP(NO_DIFF) >> PARSE(S)
(j) DROP(GENDER) >> PARSE(S)
(k) DROP(NUMBER) >> PARSE(S)
(l) DROP(PERSON) >> PARSE(S)

What is remarkable here is that DROP(NO_DIFF) outranks PARSE(S). This will be the only instance we witness in our model of DROP(NO_DIFF) ranking higher than any of the other constraints; Catalan is the only language we consider where DROP(NO_DIFF) does not reside at the bottom of the hierarchy. This reflects the ability in Catalan, demonstrated in (25ct) and (26ct) from Chapter 4, to have the subject co-referenced with either appropriate antecedent, even if the competing antecedents share identical features with each other, in two-sentence structures. Also, pair-wise constraint (i) is the only difference in our data between Catalan and Castilian; for the purposes of the present exposition, this is the minimal difference between our Castilian and Catalan EVALs.

The ranking of constraints is as follows:

DROP(PERSON) >> PARSE(PRO) >> DROP(NUMBER) >> PARSE(C) >>
{DROP(GENDER), DROP(NO_DIFF)} >> PARSE(S)

In our model, DROP(GENDER) and DROP(NO_DIFF) behave identically.
The twelve pair-wise constraint interactions for European Portuguese are as follows:

(a) PARSE(PRO) >> DROP(NO_DIFF)
(b) PARSE(PRO) >> DROP(GENDER)
(c) PARSE(PRO) >> DROP(NUMBER)
(d) DROP(PERSON) >> PARSE(PRO)

(e) PARSE(C) >> DROP(NO_DIFF)
(f) PARSE(C) >> DROP(GENDER)
(g) DROP(NUMBER) >> PARSE(C)
(h) DROP(PERSON) >> PARSE(C)

(i) PARSE(S) >> DROP(NO_DIFF)
(j) PARSE(S) >> DROP(GENDER)
(k) DROP(NUMBER) >> PARSE(S)
(l) DROP(PERSON) >> PARSE(S)

It is noteworthy that we see two significant differences between the pair-wise constraint interactions for European Portuguese and those for Brazilian Portuguese (cf. (g) and (k)). This leads us to question to what extent the two are "dialects of the same language", and whether they really should be considered as two distinct varieties. This will be treated more rigorously in the next chapter.

The ranking of constraints is as follows:

DROP(PERSON) >> PARSE(PRO) >> DROP(NUMBER) >> \{PARSE(C), PARSE(S)\} >> \{DROP(GENDER), DROP(NO_DIFF)\}

PARSE(C) and PARSE(S) act identically in our model, suggesting that in European Portuguese, the type of boundary (clausal or sentential) separating the subject and the antecedents is not as crucial as whether such a boundary is present at all. DROP(GENDER) and DROP(NO_DIFF) show no behavioural differences and reside at the bottom of the hierarchy, indicating that gender as a distinguishing feature between competing antecedents in European Portuguese (and in the Brazilian variant...
6.3.3.4 European Portuguese

The twelve pair-wise constraint interactions for European Portuguese are as follows:

(a) PARSE(PRO) >> DROP(NO_DIFF)
(b) PARSE(PRO) >> DROP(GENDER)
(c) PARSE(PRO) >> DROP(NUMBER)
(d) DROP(PERSON) >> PARSE(PRO)

(e) PARSE(C) >> DROP(NO_DIFF)
(f) PARSE(C) >> DROP(GENDER)
(g) DROP(NUMBER) >> PARSE(C)
(h) DROP(PERSON) >> PARSE(C)

(i) PARSE(S) >> DROP(NO_DIFF)
(j) PARSE(S) >> DROP(GENDER)
(k) DROP(NUMBER) >> PARSE(S)
(l) DROP(PERSON) >> PARSE(S)

It is noteworthy that we see two significant differences between the pair-wise constraint interactions for European Portuguese and those for Brazilian Portuguese (cf. (g) and (k)). This leads us to question to what extent the two are “dialects of the same language”, and whether they really should be considered as two distinct varieties. This will be treated more rigorously in the next chapter.

The ranking of constraints is as follows:

DROP(PERSON) >> PARSE(PRO) >> DROP(NUMBER) >> \{PARSE(C), PARSE(S}\} >> \{DROP(GENDER), DROP(NO_DIFF)\}

PARSE(C) and PARSE(S) act identically in our model, suggesting that in European Portuguese, the type of boundary (clausal or sentential) separating the subject and the antecedents is not as crucial as whether such a boundary is present at all. DROP(GENDER) and DROP(NO_DIFF) show no behavioural differences and reside at the bottom of the hierarchy, indicating that gender as a distinguishing feature between competing antecedents in European Portuguese (and in the Brazilian variant
as well) is rather inept at licensing pro; it may as well not be there at all, which is the situation represented by DROP(NO_DIFF).

6.3.3.5 Italian

The twelve pair-wise interactions for Italian are as follows:

(a) PARSE(PRO) >> DROP(NO_DIFF)
(b) PARSE(PRO) >> DROP(GENDER)
(c) PARSE(PRO) >> DROP(NUMBER)
(d) DROP(PERSON) >> PARSE(PRO)
(e) PARSE(C) >> DROP(NO_DIFF)
(f) PARSE(C) >> DROP(GENDER)
(g) PARSE(C) >> DROP(NUMBER)
(h) DROP(PERSON) >> PARSE(C)
(i) PARSE(S) >> DROP(NO_DIFF)
(j) PARSE(S) >> DROP(GENDER)
(k) DROP(NUMBER) >> PARSE(S)
(l) DROP(PERSON) >> PARSE(S)

Italian differs from European Portuguese only with respect to pair-wise interaction (g). The minimal difference, in our model, between the two grammars therefore is that in Italian, using the number feature to distinguish between two antecedents is not sufficient to license pro when they are separated from their potential subject co-referent by a clausal boundary. This action is possible in European Portuguese.

The ranking of constraints is as follows:

DROP(PERSON) >> {PARSE(PRO), PARSE(C)} >> DROP(NUMBER) >> PARSE(S) >> {DROP(GENDER), DROP(NO_DIFF)}

PARSE(PRO) and PARSE(C) behave identically in Italian, suggesting that separation of the subject and the antecedents by a clausal boundary in this language
does not appear to have any effect on the licensing of the null subject. Comments made for European Portuguese in Section 6.3.3.4 regarding DROP(GENDER) and DROP(NO_DIFF) are applicable here as well.

6.3.3.6 Puerto Rican Spanish

The twelve pair-wise interactions for Puerto Rican Spanish are as follows:

(a) PARSE(PRO) >> DROP(NO_DIFF)
(b) PARSE(PRO) >> DROP(GENDER)
(c) PARSE(PRO) >> DROP(NUMBER)
(d) PARSE(PRO) >> DROP(PERSON)
(e) PARSE(C) >> DROP(NO_DIFF)
(f) PARSE(C) >> DROP(GENDER)
(g) PARSE(C) >> DROP(NUMBER)
(h) PARSE(C) >> DROP(NUMBER)
(i) PARSE(S) >> DROP(NO_DIFF)
(j) PARSE(S) >> DROP(GENDER)
(k) PARSE(S) >> DROP(NUMBER)
(l) PARSE(S) >> DROP(PERSON)

We note that all three PARSE constraints outrank all four DROP constraints. This is consistent with Puerto Rican Spanish’s status as the least pro-friendly of the six Romance varieties investigated here. If all versions of PARSE are ranked higher than all versions of DROP, then we will not see any instances of the null subject in instances of co-reference with the non-topic antecedent. This can be confirmed by glancing at the even-numbered (pr) sentences in Chapter 4.

There are staggering differences in constraint behaviour between what is seen here and that for Castilian, even greater than the disparities witnessed earlier between Brazilian and European Portuguese. This is compelling evidence to examine the relation between what is commonly thought of as versions of a single language.

The ranking of constraints is as follows:

{PARSE(PRO), PARSE(C), PARSE(S)} >> {DROP(PERSON), DROP(NUMBER), DROP(GENDER), DROP(NO_DIFF)}
Based on our data, we can be no more specific than to say that the constraints in the PARSE group outrank those in the DROP set.

6.3.4 **General comments about the model**

We have just reviewed the constraint rankings of the six Romance varieties under consideration in our work. Before we return to them for cross-linguistic comparisons, it would be helpful here to resolve a number of issues about our model to strengthen its legitimacy.

6.3.4.1 **Presence of multiple features**

In Section 4.1.5, we examined instances of our data where more than one feature was present to distinguish between the two competing antecedents, and claimed that combinations of features are able to license *pro* as long as at least one feature in the combination could do so on its own. That is, a language would permit *pro* to be co-referenced with a non-topic antecedent if the antecedent was distinguishable from a competing antecedent by, for example, gender and number features, if and only if that language would permit the same if the antecedents were differentiated by just the gender feature or just the number feature alone. If neither of these two features, on its own, could tolerate *pro* in instances of switch reference, then neither would their simultaneous presence.

This may be more clearly understood through inspection of the constraint hierarchies and our knowledge of the mechanisms of Optimality Theory. The crucial interaction in resolving whether the null subject is licensed is to compare the position of the relevant PARSE constraint, determined by the syntactic structure under examination, with the relevant DROP constraints, determined by which featural differences between the two antecedents are present. In fact, by the very nature of Optimality Theory, only the highest ranked of the DROP constraints need be considered, since all the other ones below it are dominated and rendered inactive. For instance, if the highest relevant DROP constraint ranks higher than the pertinent PARSE constraint, license *pro*; whatever interaction is present between that PARSE
constraint and lower ranked DROP constraints is neutralised. If the highest relevant DROP constraint is ranked lower than the PARSE constraint in question, then so will all the other DROP constraints, and an overt subject pronoun will be required.

This is also intuitively graspable. If DROP(NUMBER) outranks PARSE(C), that indicates that the presence of number features, on its own, to distinguish between the antecedents is sufficient in the relevant syntactic environment to tolerate pro. If the two antecedents are also distinguishable by gender, there is no reason to expect that this additional differentiation would somehow undo the ability to license the null subject that was already available when number alone was used to discriminate between the antecedents.

6.3.4.2 Treating co-reference with a topic antecedent in our model

Up to this point, our framework has only addressed instances of co-reference between the subject and the non-topic antecedent. The motivation behind this was that all of the microvariation witnessed in our data from Chapter 4 occurred in instances of switch reference, whereas either pro or an overt subject pronoun was permitted whenever the subject was co-referenced with the topic antecedent, in all situations, in all languages. Therefore, we focused our model on capturing the essence of the differences that arose in the scope of co-reference with the non-topic antecedent. Now that we have presented a satisfactory model of this, it is time to explain how to account for instances of co-reference with the topic antecedent.

We see that pro is always licensed whenever co-reference occurs between the subject and the topic antecedent. Furthermore, we note that the three PARSE constraints that we have already introduced in our model only refer to co-reference with the non-topic antecedent. Therefore, whenever there is an instance of co-reference with the topic antecedent, all three PARSE constraints are vacuously satisfied. There are no remaining constraints that call for an overt subject pronoun to be actualised; the only ones left are the DROP constraints, which all call for the subject pronoun not to be realised. As a result, our model, as it is, is fully compatible with the odd-numbered examples from Chapter 4. No changes are required for this framework to be able to account for null subjects co-referenced with the topic antecedent.
6.3.4.3 Ranking the languages and features

Now that we have established six constraint hierarchies, representing the EVAL component of the grammars in each of our languages, we may now look for trends and commonalities. The first noticeable generalisation is that in all six of our ranking schemes, DROP(PERSON) is not lower than DROP(NUMBER), which in turn is not lower than DROP(GENDER). We can represent this as follows:

\[ \text{DROP(PERSON)} \geq \text{DROP(NUMBER)} \geq \text{DROP(GENDER)} \]

This can be interpreted to mean that using the person feature to disambiguate two competing antecedents will never have a lesser effect on the licensing of pro than doing so with the number feature. Similarly, one will never be worse off using the number feature to differentiate between the two competing antecedents than by using the gender feature. This is consistent with the preliminary hierarchy that we had proposed in Section 4.3. Further analysis of the differences in strength in our features will be presented in the next section.

The PARSE constraints also display a similar relationship:

\[ \text{PARSE(PRO)} \geq \text{PARSE(C)} \geq \text{PARSE(S)} \]

Note that there is tangible meaning to this order. PARSE(PRO) at the top is the situation where the antecedents are closest to the subject, whereas PARSE(S) at the lower end of the ranking corresponds to when the antecedents are furthest away from the subject\(^5\). For PARSE(C), occupying the position between the other two constraints, the distance between the subject and the antecedents is intermediate.

We also observe that the higher the PARSE constraints are in the rankings, the less amenable that language is to pro. In Puerto Rican Spanish, the three PARSE constraints are jointly ranked as highest. In Brazilian Portuguese, they are very highly positioned, being superseded only by DROP(PERSON). Italian and European

\[^5\text{This is assuming we consider separation by a sentential boundary to be greater than separation by a clausal boundary, where the subjects and the antecedents are at least in the same sentence.}\]
Portuguese have their PARSE constraints located intermittently in their rankings. Castilian has them placed lower in its hierarchy, and Catalan puts them relatively the lowest of all six Romance varieties.

We notice an inverse correlation between the position of the PARSE constraints and the level of pro-friendliness a language displays. The higher the PARSE constraints are ranked, the less it tolerates pro; Puerto Rican Spanish is a clear example of this. They are collectively ranked the lowest in Catalan, which displays the greatest number of instances of pro being licensed. We see that the placement of the PARSE constraints is thus a formalisation of the language hierarchy that we first postulated in Section 4.3, with Catalan at the top and Puerto Rican Spanish at the bottom.

This makes sense. The PARSE constraints control for the expression of an overt subject pronoun, so in languages where they are highly positioned, we expect to see fewer instances of pro. Where they are de-prioritised, this allows the DROP constraints to have greater influence, resulting in more environments where the null subject is licensed.

The constraint rankings detailed here are consistent theoretically with observations made by Aissen (1999), who asserts the existence of universally fixed constraint subhierarchies. She maintains that language-particular variation can be attributed to the “interpolation of other constraints among those in a subhierarchy, but not through differences in ranking within the subhierarchy itself” (Aissen 1999: 682). This description is indicative of what we see with our DROP and PARSE constraint rankings above, which stay fixed, but whose particular interactions with each other account for the inter-linguistic variation we witness in our six varieties.

Our results are also in tune with arguments posited by Artstein (1998). In his brief exposition on the nature of null subjects in Hebrew and other languages, he claims that the cross-linguistic variation witnessed results from the mechanism of alignment, in which multiple fixed hierarchies align with each other at various points, and it is the differing location of these meeting points for each language that explains the inter-linguistic variation (Artstein 1998: 6). This does appear to be the case with our data, where the varying merging patterns of the DROP and PARSE constraint hierarchies for each of our languages lead to the microvariation observed.

### 6.4 Unresolved issues

Now that we have developed our proposed model in adequate depth, we are equipped with stronger explanations with which to respond to issues raised in previous chapters. We now turn to addressing those questions.

#### 6.4.1 Differences in the strengths of features

It was first mentioned in Section 3.2.5 that the gender, number, and person features do not behave in an identical manner. We now have convincing evidence of this; the different positions that DROP(GENDER), DROP(NUMBER), and DROP(PERSON) occupy, both within the EVAL of any one language and also in comparisons across the different grammars as well, is testimony to their dissimilarity. Gilligan (1987: 234) foreshadows not only the realisation that gender, number, and
person identification at differing strengths, but also that the person feature is the most crucial. This is consistent with the hierarchy of DROP constraints in Section 6.3.4.3 above, where DROP(NUMBER) was positioned at the top, indicating that using number-distinguishing features to license a dropped subject pronoun was the strongest factor of the three.

Others have suggested similar arguments. Crain and Lillo-Martin (1999: 307-308) show that children learning American Sign Language develop the use of null arguments to refer first with referents present in the discourse, and only later on do they use null arguments for referents not present. This divide hints at deictic considerations in licensing null arguments in this linguistic variety, and deixis is rooted in grammatical person. Alexiadou and Anagnostopoulou (1998: 518) allude to data in Arabic to suggest that the person feature is the crucial one to the EPP.

Furthermore, it makes sense that using number is more effective than using gender to license pro because gender gets neutralised in more instances than number does. Simple verb forms generally indicate number but not gender; we can usually tell from looking at the verbal ending whether the subject is singular or plural, but not whether it is masculine or feminine. Direct clitic object pronouns normally indicate both gender and number, so neither is more clearly indicated by their usage.

It may be tempting to assert that differences in the features are grounded primarily in the morphology, but this should be cautiously considered. We note that in the languages here examined, when gender and person are used to disambiguate between the antecedents, the clitic object pronoun was, for the most part, the only clue to aid in co-referencing; the verbal ending was of little help (e.g., Puerto Rican Spanish lo odiaba identifies the object antecedent as third person and masculine, but we are not able to tell from odiaba alone whether the subject is first or third person, or masculine or feminine). Where possible, syncretic verb forms were employed so as not to indicate first, second, or third person antecedents uniquely, or masculine or feminine antecedents (distinguishment between which is generally unavailable in non-

---

6 They also mention personal communication with Anders Holmberg, who provides arguments along similar lines, pointing towards data in the historical development of Scandinavian.
7 Italian, which possesses in the imperfect and the conditional different first person and third person singular forms, provided the most instances where both clitic and verbal desinences could establish co-reference (e.g., lo odiavo indicates both that the object is masculine, and that the subject is first person singular). Note, however, that Italian did not rank at the top of our pro-drop hierarchy, either with respect to DROP(PERSON), or overall.
compound verbal forms in most Romance varieties); this was motivated by an attempt to isolate the effects of the features.

Note, however, that for grammatical number, there is no way that we can use only the clitic or only the verbal desinence to indicate co-references; both types occur together. This is because there are no verbal forms in our chosen Romance varieties where we see syncretism between singular and plural, in the same way that person and gender can be neutralised in the verbal termination. So whenever the two antecedents can be distinguished by number alone, this must always be manifested in the clitic object and the verbal ending; both pieces provide beneficial information (e.g., European Portuguese *os odi[a marks plural object and singular subject; *o odeiam marks a singular object and plural subject). Despite this apparent advantage of using the number feature to disambiguate between the competing antecedents, we see that it is bookended by both person and gender in our feature hierarchy; it is more effective than gender, but less so than person. This suggests that little influence towards licensing *pro is present in the verbal desinences alone. Similarly, both person and gender are able to disambiguate on the basis of just the clitic, yet they show vastly different strengths, so the form of the clitic itself is generally uninformative in tolerating the null subject. This advances the possibility that the differences in the strengths of the features can be ascribed not to morphological considerations alone (if at all), but must incorporate to some extent the abstract qualities of the features themselves.

6.4.2 Cole’s “semantically adequate identification”

The term “semantically adequate identification” appeared in Cole’s DROPSUBJECT constraint, and we chose not to adopt it in our proposed model. The reason for this should now be clear; we have demonstrated that semantic identification is not a unitary phenomenon because it varies according to the manner in which this is attempted. This is reflected in the differing strengths of using the three features to identify the subject to semantic adequacy.

Our approach has been to eliminate the concept of general semantic identification from the workings of the constraints, and to break this idea down into its three types, by gender, number, and person feature indication. This is manifested
compound verbal forms in most Romance varieties); this was motivated by an attempt to isolate the effects of the features.

Note, however, that for grammatical number, there is no way that we can use only the clitic or only the verbal desinence to indicate co-references; both types occur together. This is because there are no verbal forms in our chosen Romance varieties where we see syncretism between singular and plural, in the same way that person and gender can be neutralised in the verbal termination. So whenever the two antecedents can be distinguished by number alone, this must always be manifested in the clitic object and the verbal ending; both pieces provide beneficial information (e.g., European Portuguese os odiá indicates plural object and singular subject; o odeiam marks a singular object and plural subject). Despite this apparent advantage of using the number feature to disambiguate between the competing antecedents, we see that it is bookended by both person and gender in our feature hierarchy; it is more effective than gender, but less so than person. This suggests that little influence towards licensing pro is present in the verbal desinences alone. Similarly, both person and gender are able to disambiguate on the basis of just the clitic, yet they show vastly different strengths, so the form of the clitic itself is generally uninformative in tolerating the null subject. This advances the possibility that the differences in the strengths of the features can be ascribed not to morphological considerations alone (if at all), but must incorporate to some extent the abstract qualities of the features themselves.

6.4.2 Cole's "semantically adequate identification"

The term "semantically adequate identification" appeared in Cole's DROPSUBJECT constraint, and we chose not to adopt it in our proposed model. The reason for this should now be clear; we have demonstrated that semantic identification is not a unitary phenomenon because it varies according to the manner in which this is attempted. This is reflected in the differing strengths of using the three features to identify the subject to semantic adequacy.

Our approach has been to eliminate the concept of general semantic identification from the workings of the constraints, and to break this idea down into its three types, by gender, number, and person feature indication. This is manifested
in our framework by the constraints DROP(GENDER), DROP(NUMBER), and DROP(PERSON). Crucially, we have shown them not to be of equal strength; hence, grouping them together into one wide category of “semantic identification” would mask their various effects. We believe that our decision has brought us closer to a truer understanding of what constitutes semantically adequate identification because our DROP constraints provide a more rigorous explanation of it than Cole’s DROPSUBJECT, which incorporates the phrase into the definition of the constraint.

6.4.3 Cole’s “morphological maximality” and licensing mechanism of pro

Cole postulates that morphological maximality, the extent to which a verb exhibits feature agreements with its external argument, is one of two essential components, along with semantic identification, in the pro licensing mechanism, first introduced in Section 3.7.2. We are now able to understand more precisely the shortcomings of this concept. It can be observed in our data that the verb is not the only instrument used to facilitate co-referencing of the subject; the presence of a clitic pronoun can determine the object antecedent, thereby often leaving only one antecedent left to serve as the subject. This is identification by process of elimination, and although it is but an indirect means, it cannot be ignored. Furthermore, as we examined in Section 6.4.1, there are times when the verb plays a relatively passive role in identifying the antecedent because it can be co-referenced with either of the two competitors, leaving the clitic to do the duty actively. Therefore, Cole’s reliance on verbal morphology alone in his morphological maximality is flawed.

This outcome is easily predictable. We have already seen convincingly that richness of agreement is completely independent of null subjecthood. Therefore, using verbal morphology as a basis for his mechanism of null subject licensing is due to be fraught with difficulties. This, compounded with a disregard for intra-linguistic variation inherent in his mechanism, causes us to question the relevance of its role in null subject licensing.

We have no new answers either for the other unclear aspects of this concept. Cole states that for pro to be permitted, context must identify the subject at the point in which morphological maximality in that language is unable to act. That context and morphology are made discrete in this formulation is troublesome. Indeed, the two
are intertwined. We cannot say that one starts to perform as soon as the other stops. Context and morphology are crucially woven together in that we use the information encoded in the morphology as the basis by which to decide whether a certain antecedent is contextually acceptable (i.e., semantically identifiable and logical). Moreover, context and semantic identification do not work only after morphological maximality has finished its part. If a language’s morphological maximality is person and number, then it will still disallow co-reference of an antecedent with the subject if there are inconsistencies between the two in either person or number, even if they are a semantic match with regard to gender, a feature that is outside that language’s morphological maximality.

Also, we have no further insight on how morphological maximality accounts for differences between two languages at the same level, for instance, Japanese and Afrikaans. If we are forced to resort to traditional parameterisation to describe differences in null subject activity (e.g., in Japanese, context alone is strong enough to license pro, whereas in Afrikaans, it is not), then we will not have progressed past PPT analyses of null subject usage, and will continue to be burdened with its problems. Our investigation in the current chapter should convince us to move away from these approaches. A better solution lies in proposing a constraint that requires overt subject pronouns. In Afrikaans, it would have highest ranking in the grammar so as to cause fatality to a candidate whenever the constraint is violated, whereas in Japanese, it would be placed at the bottom. Admittedly, this may just be the OT version of parameterisation, but at least it does demonstrate a way that our model accounts for differences between Japanese and Afrikaans, a task unaccomplished by Cole’s mechanism. This is additional justification for doubting the effectiveness of the role of morphological maximality in the licensing of pro.

### 6.4.4 Ariel’s “accessible antecedent”

Cole refers to Ariel’s “accessible antecedent” in his pro licensing mechanism. While we are no longer convinced that his mechanism as formulated can satisfactorily explain null subject usage, the concept of antecedent accessibility is still a pertinent one and should be explored in the context of our proposed model.
Ariel mentions four factors that are relevant in determining how accessible an antecedent is: distance, competition, saliency, and unity. Our OT framework directly addresses three of these items, and touches upon the fourth. Distance is captured by the three PARSE constraints, which differ in the extent to which the antecedents are separated from the subject. Where each grammar ranks the three PARSE constraints determines how relevantly distance affects null subjecthood. In Puerto Rican Spanish, all three of them are jointly ranked, meaning that the distance separating the antecedents from the subject does not affect pro licensing patterns. On the other hand, in Castilian, where the three constraints are ranked on different levels, distance is an active factor.

Competition is reflected in the DROP constraints. If competition between antecedents exists, this means that there is no way in which features will distinguish between them. This situation is treated by DROP(NO_DIFF). If instead the other three DROP constraints are activated, this indicates that at least one of the three features is able to distinguish between the two antecedents, so only one antecedent can be co-referenced with the subject; therefore, there is no competition.

Saliency is linked to topichood. For the purposes of null subject licensing, it seems that whether an antecedent is a topic is a crucial factor because topic antecedents will always license the null subject. Having topical status means pro is permitted. However, it is not a necessary factor, because we see instances where co-reference takes place with a non-topic antecedent, and no overt subject pronoun is required. So topichood is sufficient but not mandatory; there are other situations in which co-reference between a subject and an antecedent, even if it is not a topic, can still license a null subject. These situations are determined by the interactions of the PARSE constraints and the DROP constraints in each language.

Unity is not clearly defined in Ariel’s formulation. One interpretation of the term involves discourse analysis. Since our model only consists of one- and two-sentence structures, this may not be insightful. An alternate reading is available, and has to do with our knowledge of the real world. This may prove to be more telling. We cover this in more detail with our discussion of gender-inherent predicates below.
6.4.5 Gender-inherent predicates: a second look

We return now to the dilemma created in Sections 4.4 and 4.5. This involved uncertainty in the relative ranking of European Portuguese and Italian resulting from our data on gender-inherent predicates. Up to this point in the current chapter, where only data from Sections 4.1 through 4.3 had been considered, European Portuguese had unequivocally ranked above Italian in licensing the null subject. Wherever pro was permitted in Italian, so too was it allowed in the corresponding situation in European Portuguese, while we were able to locate an instance of pro-drop in European Portuguese that was prohibited in Italian (cf. (22ep) and (22it) in Chapter 4). This had been the only instance where the two had departed from each other.

Upon examination of the data on gender-inherent predicates, however, we witnessed a situation where Italian was unequivocally more tolerant of the null subject than European Portuguese was. This created hesitations in our feature and language hierarchies as formulated in Section 4.5.

One crucial difference to note between the gender-inherent predicate sentences and the rest of the data set is that the former used agreement of the adjectival ending in the predicate (e.g., the word ‘pregnant’) to distinguish between the masculine and feminine antecedents, whereas all of our other data items used pronouns (e.g., ‘him’ or ‘her’) to disambiguate between the two. That is, in our gender-inherent predicates, the gender feature used to distinguish between antecedents was manifested by synthetic inflectional agreement, whereas in our main data, the same was accomplished via the presence of clitic agreement.

We now have a clearer understanding of Ariel’s unity condition. Our gender-inherent predicates highlight the root of the tension, the conflict between computationally encoded interpretations of where a null subject draws its antecedent (syntax), and our knowledge of the real world that dictates what can and cannot serve as the logical subject of an inherently feminine predicate (semantics and pragmatics). Our data show that Castilian, Catalan, and Italian give greater weight to semantics, allowing the null subject despite expectations that it will be co-referenced with a (masculine) topic antecedent. Brazilian Portuguese, European Portuguese, and Puerto Rican Spanish show prominence to syntax, because they use overt ‘she’ to prevent co-reference with the topic antecedent, even though we know logically that this is not possible in the real world.
Returning to our OT model, to resolve the issue of how to make compatible the higher ranking of gender indicated by clitic over gender indicated by synthetic inflectional agreement in European Portuguese, and their opposite order in Italian, into one satisfactory framework, one option is to split our DROP(GENDER) constraint into two separate constraints, according to how gender of the appropriate antecedent is marked:

DROP(GENDER_C): Drop pronoun when antecedent of subject is uniquely identified by a gender-indicating clitic

DROP(GENDER_SI): Drop pronoun when antecedent of subject is uniquely identified by gender-indicating synthetic inflectional agreement

The constraint rankings for European Portuguese and Italian are repeated below:

<table>
<thead>
<tr>
<th>European Portuguese</th>
<th>Italian</th>
</tr>
</thead>
<tbody>
<tr>
<td>DROP(PERSON)</td>
<td>DROP(PERSON)</td>
</tr>
<tr>
<td>PARSE(PRO)</td>
<td>{PARSE(PRO), PARSE(C)}</td>
</tr>
<tr>
<td>DROP(NUMBER)</td>
<td>DROP(NUMBER)</td>
</tr>
<tr>
<td>{PARSE(C), PARSE(S)}</td>
<td>PARSE(S)</td>
</tr>
<tr>
<td>{DROP(GENDER), DROP(NO_DIFF)}</td>
<td>{DROP(GENDER), DROP(NO_DIFF)}</td>
</tr>
</tbody>
</table>

Our structures testing the gender-inherent predicates consisted of two sentences, so the relevant constraint from the PARSE set is PARSE(S). Remembering that pro is permitted whenever DROP outranks PARSE and that it is forbidden when PARSE is higher than DROP, we note that in Italian, DROP(GENDER_SI) must outrank PARSE(S) in order to allow the null subject, whereas in European Portuguese, the opposite must be true. What we used to call just DROP(GENDER) is now renamed DROP(GENDER_C). Therefore, the two revised grammars now look like this:
It appears that we have now been able to reconcile the grammars of the two languages with our entire data set from Chapter 4.

One consequence of this solution is that we clearly do not have an absolute hierarchy of languages; while Catalan and Castilian still remain at the top, and Brazilian Portuguese and Puerto Rican Spanish at the bottom, we can no longer conclusively assert whether European Portuguese or Italian is ranked higher in the middle. This was perhaps to be expected eventually, particularly with languages that show intermediate tolerance for the null subject, once our proposed model expanded to an increasing number of languages and syntactic environments. So we should not be disappointed that it has occurred here. Phenomena are rarely so clear-cut and indisputable in language, and we should not take this concession as any sort of shortcoming or failure, just a realisation and acceptance of the irregular and unpredictable nature of linguistic systems.

Our general feature hierarchy is still preserved because DROP(GENDER) was positioned below both of the other DROP constraints, and there is no evidence presented here to suggest that either daughter of DROP(GENDER) must be ranked higher DROP(PERSON) or DROP(NUMBER), so we may now reformulate our first attempt from Section 6.3.4.3 as follows:

\[
\text{DROP(PERSON)} \geq \text{DROP(NUMBER)} \geq \text{DROP(GENDER}_C)/\text{DROP(GENDER}_S)I
\]

As in the previous paragraph, there is no guarantee that such a hierarchy of feature relevance will be maintained as we collect more data. This may eventually be revealed through future research.
6.4.6 Criticisms of OT syntax

While we feel that we have offered a convincing presentation in favour of treating the null subject phenomenon, and syntactic issues in general, within an Optimality Theoretic framework, there has been some opposition to this approach, and it is unlikely to cease anytime soon. Therefore, it would be fair at least to acknowledge some of these arguments existing in the literature, and the responses to them.

One critic is Newmeyer (2002), who presents three primary arguments against the approach he calls “functionally-based optimality theory (FOT)”. First, he faults FOT’s placing of the form-function interplay in the mental grammar, instead maintaining that it should arise from language acquisition and use. Second, he asserts that FOT does not conform to the standard stipulation in Optimality Theory that constraints are universal. Finally, he finds fault with two hierarchies central to FOT, the thematic and relational hierarchies, and further mentions that other hierarchies are unable to be accommodated into an acceptable framework without unnecessary complexity.

Bresnan and Aissen (2002) counter with rebuttals to Newmeyer’s claims, arguing that his conclusions are based on critical misconceptions of Optimality Theory. They tackle several of Newmeyer’s points, one of the most prominent being that it is too easy to come up with any constraint for any given linguistic phenomenon, thus reducing the effectiveness and efficiency of this explanatory model (Bresnan and Aissen 2002: 85). The pair also respond to Newmeyer’s belief that it is logically inconsistent for a constraint to be both functionally motivated and innate at the same time, addressing the latter’s criticism of the thematic and relational hierarchies (Bresnan and Aissen 2002: 89-91). Finally, they cast doubt on whether Newmeyer’s solution to the unnecessary complexity that arises with some hierarchies is, in fact, simpler and equal in its empirical coverage (Bresnan and Aissen 2002: 91-92).

We are not in a position here to attempt to offer an absolute and irrefutable resolution to this impassioned discussion. Our task has merely been to present the data and analysis in the present work as a contribution to the field of linguistic, syntactic, and OT knowledge. We call upon future researchers to incorporate our findings, however they see fit, into their work, so that they may continue this worthy debate and themselves strive towards arguing successfully their position on this question of Optimality Theoretic syntax.

In our final chapter, we explore how our language-specific constraint hierarchies affect our understanding of the languages examined here and how we categorise them. We will also recap what has been covered in the present work. This will lead us to a consideration of how our research can be continued and applied to other questions.
7.1 Summarising the present work

The present dissertation has three main theoretical goals. First, it attempts to challenge the traditional assumption that subject pronouns in the Romance languages are always optional. In most discussions focused on pro-drop, it is erroneously claimed that the usage of overt subject pronouns is never compulsory. Second, it seeks to determine whether pro-drop is a unitary phenomenon within this language family. A commonly held, yet inaccurate, belief is that the employment of null and lexical subject pronouns is uniform in each of the so-called pro-drop Romance varieties. This view is perpetuated in the literature through overly simplistic cross-linguistic comparisons of null subject activity. Third, it aims to discover whether pro-drop is best formulated as a parameter. Intra-linguistic variation and the proliferation of binary parameters that results in order to capture these language-internal differences suggest that null subjecthood is better described by language-specific orderings of violable constraints, supporting the conceptualisation of pro-drop as a scale with a range of possible values.

Chapter 1 identified the problems we hope to have addressed in this dissertation. It presented quotes from a wide range of sources, such as linguistic references, syntactic textbooks, and language-specific grammars, either implying or explicitly stating that the decision to use subject pronouns in the Romance languages is truly a free choice. Suggestions that pro-drop is assumed to be uniform in Romance are manifested through statements asserting that the way in which subject pronouns are used in a certain language is identical to that in its sister languages; some of these claims were included in this introductory chapter to illustrate this. Finally, we questioned whether the term “null subject parameter” is viable. There is mention of various metaphors that have been constructed treating pro-drop not as the customary two-way switch, but rather as a continuous cline with plenty of room for intermediate positions.

In Chapter 2, we investigated factors that, perhaps unexpectedly, are not directly relevant to pro-drop in Romance. First, we examined grammatical relations,
thematic roles, and the extent of their natural correlation. While some syntactic phenomena show clear distinctions between subject, direct object, indirect object, and oblique entities, this is not the case with the Romance null subject. Subject antecedents tend to act differently from non-subject antecedents, hinting that it is not grammatical relations, but switch reference, that is crucial. With regard to thematic roles, it would be reasonable to propose that agentive antecedents are more likely to facilitate covert subject pronouns because of their supposed compatibility with the subject position. However, the thematic role of the antecedent in fact has no effect on whether a null subject can be licensed.

We also examined whether the animacy hierarchy influences the null subject in Romance. It might have been anticipated that NPs higher on the animacy hierarchy would be ‘friendlier’ to pro because of their greater likelihood to have the thematic role AGENT. Aside from the earlier discovery that thematic roles have negligible significance on the licensing of null subjects, there are problems with linking literal animacy to the animacy hierarchy. In addition, it is not always indisputable as to which of two NPs is actually higher on the animacy hierarchy, confounding difficulties. Finally, we briefly reviewed the historical development of subject pronoun usage in French, concluding that the present obligatoriness of French subject clitic pronouns is not tied to the phonetic attrition of formerly unique verbal terminations. This discussion was brought into the context of increasing subject pronoun usage and less distinctive verbal desinences currently reported for some varieties of Caribbean Spanish, with a warning that a concrete link between these two phenomena should not, and cannot, be automatically assumed.

Chapter 3 offered a selective literature review of cross-linguistic studies on pro-drop, beginning with Gilligan’s (1987) review of null subject properties in a hundred of the world’s languages, the first and only systematic investigation on such a wide scale of this topic. This served as a good introduction to the work of Samek-Lodovici (1996), who attempts to use Optimality Theory to explain pro-drop activity in Italian and English. Building upon this initial framework, we then moved on to review Cole’s (2000) useful contribution to the field, which expands Samek-Lodovici’s investigations to a wider range of syntactic environments and greater variety of the world’s languages. Indeed, the present dissertation is largely an adaptation of Cole’s study, in that we investigate, in the same syntactic environments, the effects of grammatical person and number, in addition to Cole’s original treatment
of gender, but limit the linguistic scope of our empirical collection to six varieties within just the Romance family. Furthermore, there is a pivotal methodological difference in that Cole seeks to account for when subject pronouns are optional, preferred, or required, whereas we focus only on when they must be overt to save a structure’s grammaticality.

In Chapter 4, we presented the data collected in our investigation of the Romance null subject. There was systematic coverage, in the six Romance varieties under consideration in our work, of how overt subject pronoun requirements are affected by (1) gender, number, and person as distinguishing features between competing antecedents; (2) whether the antecedent is a topic; (3) the distance between the subject and its antecedent; and (4) gender-inherent predicates. The observations discerned from our empirical results provided us with several initial hierarchies. We were able to formulate a preliminary ranking of languages based on how pro-friendly they are; Catalan was placed at the top, followed by Castilian, European Portuguese, Italian, Brazilian Portuguese, and finally, Puerto Rican Spanish as the variety most resistant to the null subject. In addition, our data suggested an ordering of grammatical features, with person most likely to be able to license pro, followed by number, and then gender.

Chapter 5 began our attempt to find a suitable model to explain our empirical data. We first examined the ability of Principles and Parameters Theory, in both its Government and Binding and Minimalist Program formats, to accommodate our results in existing frameworks. The principal hindrance that emerged was that the language-wide, binary parameters central to PPT formulations of pro-drop are largely inadequate in accounting for the intra-linguistic variation witnessed in our data. In order to describe satisfactorily in PPT the null subject microvariation present language-internally as well as cross-linguistically, it would be necessary to proliferate additional pro-drop sub-parameters, eventually to the point of inutility. The overgeneralisation and inaccuracy that result from labeling an entire language as [+pro-drop] or [-pro-drop] in GB, or from stating that in a language T has or does not have a pronominal feature, as stipulated in MP, discouraged us from concluding that Romance pro-drop can be sufficiently explained in PPT.

This realisation led us to adopt an Optimality Theoretic approach to the Romance null subject, and this was the principal focus of Chapter 6. The OT framework was made compatible with our empirical data by splitting Samek-
of gender, but limit the linguistic scope of our empirical collection to six varieties within just the Romance family. Furthermore, there is a pivotal methodological difference in that Cole seeks to account for when subject pronouns are optional, preferred, or required, whereas we focus only on when they must be overt to save a structure’s grammaticality.

In Chapter 4, we presented the data collected in our investigation of the Romance null subject. There was systematic coverage, in the six Romance varieties under consideration in our work, of how overt subject pronoun requirements are affected by (1) gender, number, and person as distinguishing features between competing antecedents; (2) whether the antecedent is a topic; (3) the distance between the subject and its antecedent; and (4) gender-inherent predicates. The observations discerned from our empirical results provided us with several initial hierarchies. We were able to formulate a preliminary ranking of languages based on how pro-friendly they are; Catalan was placed at the top, followed by Castilian, European Portuguese, Italian, Brazilian Portuguese, and finally, Puerto Rican Spanish as the variety most resistant to the null subject. In addition, our data suggested an ordering of grammatical features, with person most likely to be able to license pro, followed by number, and then gender.

Chapter 5 began our attempt to find a suitable model to explain our empirical data. We first examined the ability of Principles and Parameters Theory, in both its Government and Binding and Minimalist Program formats, to accommodate our results in existing frameworks. The principal hindrance that emerged was that the language-wide, binary parameters central to PPT formulations of pro-drop are largely inadequate in accounting for the intra-linguistic variation witnessed in our data. In order to describe satisfactorily in PPT the null subject microvariation present language-internally as well as cross-linguistically, it would be necessary to proliferate additional pro-drop sub-parameters, eventually to the point of inutility. The overgeneralisation and inaccuracy that result from labeling an entire language as [+pro-drop] or [-pro-drop] in GB, or from stating that in a language T has or does not have a pronominal feature, as stipulated in MP, discouraged us from concluding that Romance pro-drop can be sufficiently explained in PPT.

This realisation led us to adopt an Optimality Theoretic approach to the Romance null subject, and this was the principal focus of Chapter 6. The OT framework was made compatible with our empirical data by splitting Samek-
Lodovici’s and Cole’s initial DROP and PARSE constraints into their more specific versions, which capture better the effects of syntactic distance and grammatical features on pro-drop activity in the languages investigated. We also demonstrated that there are substantial benefits to treating pro-drop in an OT framework instead of PPT. These include: 1) capturing more precisely the intra-linguistic variation observed in null subject usage; 2) facilitating cross-linguistic comparisons in pro-drop activity by using the position of the PARSE constraints as a measure of a language’s friendliness to pro; and 3) a concise simplicity of exposition that is not available in PPT, where the creation of an almost endless number of sub-parameters to explain the same data make accounting for the Romance null subject overly complex, impractical, and unintuitive.

The investigation attempted in this dissertation has produced some tangible findings relating to Romance pro-drop. We now outline below some of the principal original contributions to our understanding of the null subject phenomenon.

### 7.2 Differences, groupings, and constraint hierarchies: some new insights

#### 7.2.1 Determining the identity of each language

We have expended most of our efforts in the present study on using the constraints to account for the intra-linguistic microvariation observed in our data; this was addressed in considerable depth in the previous chapter. We now turn here to cross-linguistic comparisons.

The particular ordering of our constraints in the EVAL of each language’s grammar can pinpoint the precise differences among our languages. We repeat the constraint rankings that we have established for all six Romance varieties below:\(^1\)

---

\(^1\) The behaviour of Brazilian Portuguese, Castilian, Catalan, and Puerto Rican Spanish with regard to gender-inherent predicates is consistent with what would be predicated by the hierarchies that had been established before we detailed the split of DROP(GENDER) into DROP(GENDER_C) and DROP(GENDER_SI) in Section 6.4.5. Therefore, to facilitate presentation, we keep the constraint \(\text{DROP(GENDER)}\) in those grammars, with the understanding that they may just as validly be substituted in the same place in the hierarchies by the constraint pair \(\{\text{DROP(GENDER_C)}, \text{DROP(GENDER_SI)}\}\), with no difference in end result. We maintain the two separate gender constraints in the grammars of European Portuguese and Italian.
<table>
<thead>
<tr>
<th>Brazilian Portuguese</th>
<th>Castilian</th>
<th>Catalan</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{DROP(PERSON)})</td>
<td>(\text{DROP(PERSON)})</td>
<td>(\text{DROP(PERSON)})</td>
</tr>
<tr>
<td>({\text{PARSE(PRO), PARSE(C), PARSE(S)}})</td>
<td>(\text{PARSE(PRO)})</td>
<td>(\text{PARSE(PRO)})</td>
</tr>
<tr>
<td>({\text{DROP(NUMBER), DROP(GENDER), DROP(NO_DIFF)}})</td>
<td>({\text{DROP(NUMBER), DROP(GENDER), DROP(NO_DIFF)}})</td>
<td>({\text{DROP(GENDER), DROP(NO_DIFF)}})</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>European Portuguese</th>
<th>Italian</th>
<th>Puerto Rican Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{DROP(PERSON)})</td>
<td>(\text{DROP(PERSON)})</td>
<td>{\text{PARSE(PRO), PARSE(C), PARSE(S)}})</td>
</tr>
<tr>
<td>(\text{PARSE(PRO)})</td>
<td>({\text{PARSE(PRO), PARSE(C)}})</td>
<td>{\text{PARSE(PRO), PARSE(C), PARSE(S)}})</td>
</tr>
<tr>
<td>(\text{DROP(NUMBER)})</td>
<td>({\text{DROP(NUMBER), DROP(GENDER_SI)}})</td>
<td>{\text{DROP(PERSON), DROP(NUMBER), DROP(GENDER), DROP(NO_DIFF)}})</td>
</tr>
<tr>
<td>({\text{PARSE(C), PARSE(S)}})</td>
<td>(\text{PARSE(S)})</td>
<td>{\text{PARSE(PRO), PARSE(C), PARSE(S)}})</td>
</tr>
<tr>
<td>({\text{DROP(GENDER_C), DROP(GENDER_SI), DROP(NO_DIFF)}})</td>
<td>({\text{DROP(GENDER_C), DROP(NO_DIFF)}})</td>
<td>{\text{DROP(PERSON), DROP(NUMBER), DROP(GENDER), DROP(NO_DIFF)}})</td>
</tr>
</tbody>
</table>

The set of rankings above clearly answer the questions that had been asked at the beginning of the present work. First, it is indisputable that subject pronouns are not always optional, in any of the six varieties studied; this is readily discerned by the observation that in none of the varieties do all of the DROP constraints outrank all of the PARSE constraints. Second, the juxtaposition above makes explicit the fact that the Romance languages are not homogeneous with respect to the licensing of the null subject, and that pro-drop is certainly not a unitary phenomenon within this language family. Finally, that we are offering these constraint rankings, instead of a system of sub-parameters, as the optimal model of capturing Romance null subject tendencies is testimony to our belief that pro-drop is better explained as a cline instead of a series of binary parameters.

Each of these varieties may now be identified, uniquely, by highlighting how their constraints are ranked. We may say, for example, when asked to identify characteristics particular to Castilian, that it is a language where the constraints are ranked precisely as follows:
DROP(PERSON) >> PARSE(PRO) >> DROP(NUMBER) >> PARSE(C) >> DROP(GENDER) >> PARSE(S) >> DROP(NO_DIFF)

No other language examined here shows this exact order; it is specific solely to Castilian. The same process of stating the unique rankings of the constraints applies if we are seeking to single out any of the other Romance varieties as well.

We may also try to predict the constraint rankings of those languages not investigated in the present work. For instance, if we accept the traditional view that French is a non-null subject language, then we should propose, like in Afrikaans, that a constraint requiring an overt subject pronoun be placed at the top of its hierarchy in EVAL. The present void of native speakers of Latin makes any attempt to decipher its null subject licensing grammar merely speculative, but if we were to hazard such a guess, then we might cite, among other notables, the relative few occurrences of discourse participant pronouns and the non-existence of proper third person subject pronouns (Vincent 1988a: 42), and suspect that the DROP constraints ranked quite high in the Latin hierarchy, whereas the PARSE constraints were positioned rather low. The best method of determining the actual EVAL systems of languages not covered in this research, of course, is to test our sentences in Occitan, Romanian, Sardinian, Galician, and other sister varieties. Furthermore, the analyses and models developed in this chapter are not exclusive to Romance; they are transferable to other language groups as well.

7.2.2 Pinpointing differences

Defining the terms "language" and "dialect" is tricky; there is no shortage in the literature of the difficulties that arise when one attempts this task (e.g., Chambers and Trudgill (1980: 5), Crystal (2000: 7-9)), and certainly no paucity of controversies and disagreements that result (e.g., Allen and Linn (1986), van Leuvensteijn and Berns (1992), Mattheier (2000)). We certainly do not venture a resolution here. Whatever position one does take on the issue, there can be no disputing the fact that the six Romance varieties examined in this work are autonomous linguistic systems, each separate in its own right. One only need look at the constraint rankings present
in each variety’s grammar to appreciate this, and these constraints represent only a tiny fraction of a full study of a single particular phenomenon, null subjecthood.

If there are questions as to whether Brazilian and European are two versions of the same language called Portuguese, then all it takes is a quick glimpse of their respective constraint rankings. The former has all three PARSE constraints jointly ranked, splitting DROP(PERSON) from the other three DROP constraints. The latter ranks PARSE(PRO) higher than its two cousins, and places DROP(NUMBER) in a different position. These are just two points of divergence. Our hierarchies actually support the view that European Portuguese has more in common with Italian than with Brazilian Portuguese, at least with respect to null subjects. To identify one of a number of significant differences, we can mention that in European Portuguese, DROP(NUMBER) outranks both PARSE(C) and PARSE(S), whereas the exact opposite holds in Brazilian Portuguese.

The disparities become even more pronounced when we examine Castilian and Puerto Rican Spanish. The former has all seven constraints separately ranked, one above the other, whereas in the latter, there are just two distinct tiers, the PARSE band outranking the DROP set. Relatively few of their corresponding items of data from Chapter 4 match; this is reflected in the observation that their constraint rankings hardly correlate. It does seem that we are talking about two widely disparate grammars here.

Using our model to perform cross-linguistic comparisons is obviously not limited to instances of “language versus dialect” discussions; we may also conduct them between incontestably distinct languages. If we are looking to ascertain unambiguous differences between, say, Catalan and Italian, we can simply state that in Catalan, PARSE(PRO) is ranked above PARSE(C) and PARSE(S) resides at the bottom of the hierarchy, whereas in Italian, PARSE(PRO) and PARSE(C) are jointly positioned and that PARSE(S) is one step up from the bottom of the constraint rankings. As the constraints, and their rankings, constitute each language’s EVAL, and hence determine the essence of their grammars, we can merely allude to those constraints, and their differences in positioning, to detect fundamental differences between their linguistic systems.

It would be worth looking at these results from a historical aspect as well. If we are to assume that Latin was quite tolerant of the null subject and that in its
grammar, the DROP constraints far outranked the PARSE constraints\(^2\), then we could hypothesise that Catalan is the most conservative of the six Romance varieties we have studied in this work, having the highest collective ranking of DROP constraints in its grammar. Alternatively, Puerto Rican Spanish would appear to be the most innovative of the six, because its grammar shows unequivocal ranking of the PARSE constraints over the DROP constraints, a complete inversion of what we might expect for Latin.

Another relevant question is whether our framework is suited for diachronic investigations of language. Soliciting grammatical judgments of, for example, 18\(^{th}\) century Catalan is of course now impossible, and this restriction would be a serious, if not fatal, impediment to its analysis in our model. However, if we were to take the view that, in the current absence of native speakers of any historical variety of a language, data collected from prose written in that era, and, in particular, any natural dialogues contained therein, constituted a sufficiently accurate (albeit admittedly imperfect) base to work with, this obstacle could be partly overcome. Furthermore, this research might provide us with additional clues on which of our constraints are historically the most mobile, thus contributing to our understanding of how languages change over time.

Perhaps a better solution lies in testing modern Sardinian in our framework. Commonly considered to be the most conservative Romance language presently spoken (Harris 1988: 20; Jones 1988: 314), in part due to strong resemblances in its verbal system to that of Latin, Sardinian may prove to be the most insightful regarding how Latin would be treated in our model. An investigation of Sardinian might be able to answer a number of questions posed here. For instance, does it in fact show higher tolerance for the null subject than Catalan? If not, what does that say about Sardinian’s apparent striking similarity to Latin, or our postulation of the constraint rankings in Latin? If, as we have argued earlier, verbal morphology has nothing to do with the null subject phenomenon, should we even expect Sardinian to predict the case for Latin any more accurately than the other Romance languages?

\(^2\) Again, due to the lack of appropriate data to confirm this, our assumption can only be an educated guess at best.
Thinking about traditional alignments

The data, the constraint rankings, and the differences among them lead us to consider the validity and effectiveness of traditional geographical classifications. Is it still reasonable to divide languages according to the region in which they are spoken?

To a considerable extent, our analysis does support the status quo. A legitimate argument may be made that the distinction between Ibero-Romance and the rest of the Romance family is maintained. We did observe that Catalan, Castilian, and (for the most part) European Portuguese occupied the top places in our language hierarchy of pro-drop tolerance. Looking further into Continental Ibero-Romance, we see that our language hierarchy is consistent with two divisions that are routinely postulated. The first is geographical; we have Catalan and Castilian, the eastern Ibero-Romance varieties, both grouped above European Portuguese, a western Ibero-Romance language, on our hierarchy. Examination of the null subject properties of another western Ibero-Romance variety, such as Galician, could support or dispel this split. The second break is historical; Harris asserts that “within the Iberian peninsula, the major early division, apparent [...] as early as the ninth century, was between Catalan on the one hand [...] and the other dialects of Spain and Portugal, collectively referred to as Hispano-Romance” (1988: 6). This too is reflected in our pro-drop hierarchy, with both Castilian and European Portuguese positioned below Catalan. So the hierarchy proposed here is compatible with previous claims that Castilian appears to occupy a geographically and linguistically intermediate position between Catalan and European Portuguese on the Ibero-Romance dialect continuum.

In the absence of sufficient data, we may also say that divisions between continental western Romance and eastern Romance are still preserved. Catalan, Castilian, and European Portuguese can be clumped together in certain respects, leaving Italian on its own. Investigation of, for example, Galician and Asturian in the western group, and Romanian and the numerous linguistic varieties spoken in Italy for the eastern group, may move towards either confirming or rejecting this East-West chasm.

Our six Romance varieties may also be considered roughly in terms of hub-and-spoke. We have a central area, the Iberian peninsula, and a periphery radiating away from it, Italian being the closest, then the Caribbean and South American varieties. Again, our analysis shows no reason to disregard this approach. We do see
Brazilian Portuguese and Puerto Rican Spanish showing several similarities to each other as well as significant differences from the four varieties spoken on the other side of the Atlantic. Italian, not so removed from the core, does not depart from the Iberian languages as much as Brazilian Portuguese and Puerto Rican Spanish do.

None of this can be considered absolute and irrefutable, of course. We currently do not have any reason to reject this wave model of language relationships, or any other conceptualisation based on geography, but we are also in no position to assert conclusively that this must definitely be the case. The sampling of languages investigated in this work is too small to be telling, their regional distribution can hardly be considered representative, much less all-encompassing, and the depth into which we explore each language is too shallow, for this supposition to be anything more than merely indicative and suggestive at the present moment. Future research will be able to provide us with more relevant information, and perhaps a more decisive answer.

7.3 Areas of future research

The current investigation has revealed much new information on how pro-drop works in the six Romance varieties we examined. However, there is plenty more to be discovered. We now take a brief look at how our research may serve as a suitable point of departure towards answering some other linguistic uncertainties.

7.3.1 More languages

A more thorough understanding of Romance null subjecthood, we claim, can be obtained by repeating the investigation undertaken in this dissertation in other sister languages. There is a plethora of varieties that still remain to be examined. We have numerous Latin American and Caribbean varieties of Spanish, as well as those versions spoken by immigrant communities in the United States, available for testing. This may give us particularly insightful information on whether the expatriate variety of, for example, Guatemalan Spanish, has diverged from the home variety, or whether
it has been largely preserved in its new spoken environment. Lusophonic data may be collected in the former Portuguese colonies of Africa, such as Angola, Mozambique, and Guinea-Bissau, to see how they compare with each other and with Brazilian and European Portuguese. There is considerable variation attested in the Catalan spoken in Catalonia, the Balearic Islands, Valencia, Alghero on the Italian island of Sardinia, the southwest of France, and Andorra (Harris 1988: 12-13; Wheeler 1988: 170). Will they show differences in their null subject properties, and if so, can this be attributed at all to the influences of the different national and official languages also present?

To extend this line of reasoning, we have a number of minority languages spoken in Spain, such as Asturian and Galician, to name just two. Will the Castilian spoken by Asturians and Galicians display different null subject rules from that spoken by Basques, Catalans, or Madrileños? Will pro-drop tendencies in the Italian spoken by those of the North, Central, Upper South, and Extreme South of the nation show similarities or heterogeneity? Standard Italian, more so than any other Romance language investigated in this work (cf. the situation in Catalan (Wheeler 1988: 207)), is significantly influenced by dialectal and regional varieties\(^3\), and as such it "evinces less homogeneity than any other Romance vernacular that has achieved the status of a national language" (Vincent 1988b: 279). Factors affecting various speakers' judgments of null subject usage in standard Italian might include related phenomena in their local dialects, such as subject clitics in Northern Italian dialects and the apparent breakdown of pro-drop in Ripiano once person is lost in favour of gender (Ledgeway 2004). Of course, in addition to these discoveries, there is great intrinsic worth and interest in the examination of the null subject phenomenon in Asturian, Galician, and the various dialects of Italy as well.

Furthermore, our framework is, of course, not exclusive to Romance; it may be suitably applied to other language families as well. For instance, the Slavic languages are often claimed to be members of the pro-drop community. However, the same questions we asked at the beginning of our work about the Romance languages are equally valid for this linguistic family. Is it always the case that subject pronouns are optional in Slavic? Is pro-drop in Slavic a unitary phenomenon? With regard to the latter query, there seems to be ample scope for disagreement. For

\(^3\) By these, we largely mean the sister languages of standard Italian (i.e., the variety claiming Tuscan as its source (Giannelli (1997: 297))), such as Calabrian, Abruzzese, and Piemontese.
example, Fowler and Choo (2003) attest to the “striking uniformity in pro-drop usage across the Slavic languages”, whereas Caflisch (1997) proposes a pro-drop continuum in languages such as Czech, Polish, Russian, and what are now Serbian and Croatian. The collection of empirical data in these languages to mirror what we have attempted in Romance here may provide more concrete answers to these debates.

7.3.2 More linguistic environments

It is acknowledged that the linguistic environments under investigation in this work constitute only a minute proportion of all the factors that could affect null subject usage; this limitation is due to necessary restrictions on the time and scope of the present research, not in their level of interest or relevance. An expansion of the morphosyntactic conditions tested would contribute much welcomed insight into our understanding of the null subject phenomenon.

Samek-Lodovici addressed postverbal and expletive subjects in his model, two aspects that we were unable to treat in our work. Integrating these considerations back into our expanded framework may provide a clearer picture of how referential and expletive subjects differ, within each language and across languages. An exciting consequence of this is that it may allow us to incorporate into our study the Germanic languages, a group traditionally excluded from discussions on the null subject, despite their tolerance, in some cases, for null expletive subjects. Indeed, this may be a step in the right direction in bringing into the scope of our framework those languages that show only marginal null subject activity.

There is a role for discourse analysis in this investigation. Our study has been limited to one- and two-sentence structures, and the greatest distance between the subject and antecedent in our work is just one sentential boundary. We can easily increase the distance between the subject and antecedent by expanding the domain of linguistic environments from single- and dual-sentence structures to multi-sentence paragraphs. This would allow us to test how much the ability to co-reference a subject with its antecedent changes as a function of the syntactic and temporal distance between them. Cole (2000: 223-225) does in fact allude to this briefly in his quick exposition on paragraph-internal co-referencing; a more systematic and thorough treatment of this awaits, and would be worth pursuing.

242
The present research has been concerned primarily with syntactic considerations, although there has been some attempt to amalgamate the role of semantics into our framework, such as the discussion on gender-inherent predicates. This could be extended beyond gender to non-binary and less clearly definable characteristics, such as age. Testing age, in fact, would present an intriguing foil to how we tested gender. While we can indicate gender as either ‘masculine’ or ‘feminine’, we cannot readily do so with age by producing the labels ‘old’ and ‘young’. Indeed, this cannot be easily accomplished if we wish to formalise age as a grammatical feature just like gender, number, and person. For instance, if we were to test the sequence, ‘John is taking his grandfather to the city centre. (He) needs to pick up his pension cheque for the month.’, the semantics make it rather clear who the intended subject of the second sentence is. However, there is no formal marking of [+old] on ‘grandfather’, or [-old] on ‘John’, in the same way that Mary was [+feminine] and John was [+masculine] in our gender-inherent predicates. Thus, this would bring up two possible questions: (1) whether age as a semantic cue for subjecthood works in the same way as gender, and (2) whether semantically suggestive predicates work the same way for formal grammatical features, like gender, as for those that cannot be strictly formalised, like age.

Finally, all of our interest in pro-drop has been limited to the subject position. Of course, the occurrence of pro is not restricted to grammatical subjects; it may occur in other argument positions as well. Rizzi (1986) and Raposo (1986) discuss direct object pro in Italian and European Portuguese, respectively, in great detail, and others have cited oblique pro in many Germanic languages (e.g., ‘Are you coming with?’). Most recently, Cummins and Roberge (2004) discuss null objects in French, a language often conspicuous by its absence from discussions of this area of Romance syntax. Furthermore, null objects in French apparently must be inanimate; this contrasts with our discovery in Chapter 2 that animacy does not have a direct effect on subject pro licensing in Romance. This, then, would represent a propitious opportunity to end this language’s segregation and integrate it back into comparative pan-Romance investigations.

There is no reason why the framework used in the present research cannot be applied to these non-subject instances of pro. This is particularly true for those languages that indicate object agreement on its verbal desinences. It would be telling to see whether, in these languages, antecedents for the null object compete in a similar
manner, with respect to topichood and distinguishing grammatical features, as the antecedents for subjects in our study.

The field of Romance linguistics is a rich and varied one. It is hoped that the data, results, and conclusions presented in this dissertation will constitute a tangible contribution to this area of study, however limited in scope our offering may be. Furthermore, this work will have been considered a success if it has been able to draw much-needed attention to the vast wealth of tradition and scholarship in Romance philology, and to the continued need to keep exploring the innumerable phenomena still waiting to be investigated, in the multitude of Romance varieties in existence today. Finally, our research has concentrated on the considerable microvariation witnessed within this closely related family of languages, and so it attempts to encourage an appreciation for linguistic diversity in an era where English is rapidly becoming a default lingua franca worldwide.
APPENDIX A

BRAZILIAN PORTUGUESE

AGENT

1. O Pedro leu as poesias no funeral. *(Ele) falou com muita eloqüência.
2. Fizeram falar o Pedro no funeral. *(Ele) falou com muita eloqüência.
3. Fizeram ler as poesias ao Pedro no funeral. *(Ele) falou com muita eloqüência.
4. As poesias foram lidas pelo Pedro. *(Ele) falou com muita eloqüência.

THEME

5. O Miguel foi preso. *(Ele) tentou escapar mas sem êxito.
6. Contiveram o Miguel. *(Ele) tentou escapar mas sem êxito.
7. Resistiram ao Miguel. *(Ele) tentou avançar mas sem êxito.
8. Tomaram conta do Miguel. *(Ele) tinha comido uma omelete o dia antes e tinha adoecido.

EXPERIENCER

9. O Marco tem medo dos cães. *(Ele) chora se os vê.
10. Os cães assustam o Marco. *(Ele) chora se os vê.
11. Os cães são antipáticos ao Marco. *(Ele) chora se os vê.
12. Para o Marco, os cães deveriam estar atados. *(Ele) chora se os vê.

BENEFACTIVE

13. O João foi ajudado graças a algumas doações. *(Ele) pode pagar agora a operação.
15. Aplicaram ao João as regras acerca dos pagamentos de deficiência. *(Ele) pode pagar agora a operação.

16. Colecionaram muito dinheiro para o João. *(Ele) pode pagar agora a operação.

RECIPIENT

17. O Paulo recebeu muitos postais dos seus amigos. *(Ele) agradeceu a todos pelos pensamentos amáveis deles.

18. Mimaram o Paulo de afeto. *(Ele) agradeceu a todos pela bondade deles.

19. Daram muitos presentes ao Paulo. *(Ele) agradeceu a todos pela generosidade deles.


CASTILIAN SPANISH

AGENT

1. Pedro leyó las poesías en el funeral. *(El) habló con mucha elocuencia.

2. Hicieron hablar a Pedro en el funeral. *(El) habló con mucha elocuencia.

3. Le hicieron leer las poesías a Pedro en el funeral. *(El) habló con mucha elocuencia.

4. Las poesías fueron leidas en el funeral por Pedro. *(El) habló con mucha elocuencia.

THEME

5. Miguel fue arrestado. *(El) trató de escapar pero sin éxito.

6. Refrenaron a Miguel. *(El) trató de escapar pero sin éxito.

7. Le resistieron a Miguel. *(El) trató de avanzar pero sin éxito.

8. Cuidaron de Miguel. *(El) había comido una tortilla el día anterior y se había puesto enfermo.
EXPERIENCER


10. Los perros espantan a *Marco*. *(El)* llora se los ve.

11. Los perros le son antipáticos a *Marco*. *(El)* llora se los ve.

12. Para *Marco* los perros deberían estar atados. *(El)* llora se los ve.

BENEFACTIVE

13. *Juan* fue ayudado gracias a algunos donativos. *(El)* puede pagar ahora la operación.

14. Ayudaron a *Juan* con mucho auxilio financiero. *(El)* puede pagar ahora la operación.

15. Le aplicaron a *Juan* las reglas sobre los pagos de incapacidad. *(El)* puede pagar ahora la operación.


RECIPIENT

17. *Pablo* recibió muchas postales de sus amigos. *(El)* les agradeció sus pensamientos amables a todos.

18. *Mimaron* a *Pablo* con cariño. *(El)* les agradeció su bondad a todos.

19. Le dieron muchos regalos a *Pablo*. *(El)* les agradeció su generosidad a todos.

20. *Oraron* en nombre de *Pablo*. *(El)* les agradeció su preocupación a todos.

CATALAN

AGENT


2. *Van fer parlar* en *Pere* al funeral. *(El)* va parlar amb molta eloquència.

3. *Van fer llegir les poesies a en Pere* al funeral. *(El)* va parlar amb molta eloquència.
4. Les poesies van ser llegides per en Pere. (Els) va parlar amb molta eloqüència.

**THEME**

5. En Miquel va ser arrestat. (Els) va intentar de escapar però sense l’èxit.
8. Van tenir cura d’en Miquel. (Els) havia menyat una truita el dia abans i havia emmalaltit.

**EXPERIENCER**

9. En Marc té por dels gossos. (Els) crida si els veu.
10. Els gossos espanten en Marc. (Els) crida si els veu.
11. Els gossos són antipàtics a en Marc. (Els) crida si els veu.
12. Per en Marc, els gossos haurien de estar lligats. (Els) crida si els veu.

**BENEFACTIVE**

15. Van aplicar a en Joan les regles sobre els pagaments de incapacitat. (Els) pot pagar ara l’operació.
16. Van recaptar molts diners per en Joan. (Els) pot pagar ara l’operació.

**RECIPIENT**

17. En Pau va rebre moltes targetes dels seus amics. (Els) va regraciar tots per els seus pensaments amables.
18. Van aviciar en Pau de afectió. (Els) va regraciar tots per la seva bondat.
19. Van donar molts regals a en Pau. (Els) va regraciar tots per la seva generositat.
20. Van pregar en nom d’en Pau. (Ell) va regraciar tots per la seva preocupació.

EUROPEAN PORTUGUESE

AGENT

1. O Pedro leu as poesias no funeral. (Ele) falou com muita eloquência.

2. Fizeram falar o Pedro no funeral. (Ele) falou com muita eloquência.

3. Fizeram ler as poesias ao Pedro no funeral. (Ele) falou com muita eloquência.

4. As poesias foram lidas pelo Pedro. (Ele) falou com muita eloquência.

THEME

5. O Miguel foi preso. (Ele) tentou escapar mas sem êxito.

6. Contiveram o Miguel. (Ele) tentou escapar mas sem êxito.

7. Resistiram ao Miguel. (Ele) tentou avançar mas sem êxito.

8. Tomaram conta do Miguel. (Ele) tinha comido uma omelete o dia antes e tinha adoecido.

EXPERIENCER

9. O Marco tem medo dos cães. (Ele) chora se os vê.

10. Os cães assustam o Marco. (Ele) chora se os vê.

11. Os cães são antipáticos ao Marco. (Ele) chora se os vê.

12. Para o Marco, os cães deveriam estar atados. (Ele) chora se os vê.

BENEFACTIVE

13. O João foi ajudado graças a algumas doações. (Ele) pode pagar agora a operação.


15. Aplicaram ao João as regras acerca dos pagamentos de deficiência. (Ele) pode pagar agora a operação.
16. Coleccionaram muito dinheiro para o João. (Ele) pode pagar agora a operação.

**RECIPIENT**

17. O Paulo recebeu muitos postais dos seus amigos. (Ele) agradeceu a todos pelos pensamentos amáveis deles.

18. Mimaram o Paulo de afecto. (Ele) agradeceu a todos pela bondade deles.

19. Daram muitos presentes ao Paulo. (Ele) agradeceu a todos pela generosidade deles.


**ITALIAN**

**AGENT**

1. Pietro ha letto le poesie al funerale. (Lui) ha parlato con molta eloquenza.

2. Hanno fatto parlare Pietro al funerale. (Lui) ha parlato con molta eloquenza.

3. Hanno fatto leggere le poesie a Pietro al funerale. (Lui) ha parlato con molta eloquenza.

4. Le poesie sono state lette da Pietro al funerale. (Lui) ha parlato con molta eloquenza.

**THEME**

5. Michele è stato arrestato. (Lui) ha tentato di scappare ma senza successo.

6. Hanno trattenuto Michele. (Lui) ha tentato di scappare ma senza successo.

7. Hanno resistito a Michele. (Lui) ha tentato di avanzare ma senza successo.

8. Hanno preso cura di Michele. (Lui) aveva mangiato una frittata il giorno prima e s'era ammalato.

**EXPERIENCER**

9. Marco ha paura dei cani. (Lui) piange se li vede.
10. I cani spaventano Marco. (Lui) piange se li vede.

11. I cani sono antipatici a Marco. (Lui) piange se li vede.

12. Per Marco i cani dovrebbero stare legati. (Lui) piange se li vede.

**BENEFACTIVE**


14. Hanno aiutato Gianni con molta assistenza finanziaria. (Lui) può pagare adesso l’operazione.

15. Hanno applicato a Gianni le regole riguardo ai pagamenti di incapacità. (Lui) può pagare adesso l’operazione.

16. Hanno raccolto molti soldi per Gianni. (Lui) può pagare adesso l’operazione.

**RECIPIENT**

17. Paolo ha ricevuto molte cartoline dai suoi amici. (Lui) ha ringraziato tutti per i loro gentili pensieri.

18. Hanno riempito Paolo di affetto. (Lui) ha ringraziato tutti per il loro gentilezza.

19. Hanno dato molti regali a Paolo. (Lui) ha ringraziato tutti per il loro generosità.

20. Hanno pregato per conto di Paolo. (Lui) ha ringraziato tutti per il loro interesse.

**PUERTO RICAN SPANISH**

**AGENT**

1. Pedro leyó las poesías en el funeral. (El) habló con mucha elocuencia.

2. Hicieron hablar a Pedro en el funeral. *(El) habló con mucha elocuencia.

3. Le hicieron leer las poesías a Pedro en el funeral. *(El) habló con mucha elocuencia.

4. Las poesías fueron leídas en el funeral por Pedro. *(El) habló con mucha elocuencia.
THEME
5. Miguel fue arrestado. *(EI) trató de escapar pero sin éxito.
6. Refrenaron a Miguel. *(EI) trató de escapar pero sin éxito.
7. Le resistieron a Miguel. *(EI) trató de avanzar pero sin éxito.
8. Cuidaron de Miguel. *(EI) había comido una tortilla el día anterior y se había puesto enfermo.

EXPERIENCER
9. Marco tiene miedo de los perros. *(EI) llora se los ve.
10. Los perros espantan a Marco. *(EI) llora se los ve.
11. Los perros le son antipáticos a Marco. *(EI) llora se los ve.
12. Para Marco los perros deberían estar atados. *(EI) llora se los ve.

BENEFACTIVE
13. Juan fue ayudado gracias a algunos donativos. *(EI) puede pagar ahora la operación.
14. Ayudaron a Juan con mucho auxilio financiero. *(EI) puede pagar ahora la operación.
15. Le aplicaron a Juan las reglas sobre los pagos de incapacidad. *(EI) puede pagar ahora la operación.
16. Recaudaron mucho dinero para Juan. *(EI) puede pagar ahora la operación.

RECIPIENT
17. Pablo recibió muchas postales de sus amigos. *(EI) les agradeció sus pensamientos amables a todos.
18. Mimaron a Pablo con cariño. *(EI) les agradeció su bondad a todos.
19. Le dieron muchos regalos a Pablo. *(EI) les agradeció su generosidad a todos.
20. Oraron en nombre de Pablo. *(EI) les agradeció su preocupación a todos.
APPENDIX B

BRAZILIAN PORTUGUESE

1. Toda responsabilidade fica comigo. (Eu) tenho de aceitar as consequências.

4. Toda responsabilidade fica com o João S. (Eu) tenho de aceitar as consequências.

7. Toda responsabilidade fica com o autor presente. (Eu) tenho de aceitar as consequências.

10. Toda responsabilidade fica com você. *(Você) tem de aceitar as consequências.

13. Toda responsabilidade fica com João S. *(Você) tem de aceitar as consequências.

16. Toda responsabilidade fica com o médico. *(Você) tem de aceitar as consequências.

19. Toda responsabilidade fica com ele. *(Ele) tem de aceitar as consequências.

20. A criança está chorando por causa de aquele. *(Ele) a assustou.

21. A criança está chorando por causa de aquela. *(Aquela) a assustou.

22. Toda responsabilidade fica com o João S. *(Ele) tem de aceitar as consequências.

23. A criança está chorando por causa do Rex. *(Ele) a assustou.


25. Toda responsabilidade fica com o menino. *(Ele) tem de aceitar as consequências.

26. A criança está chorando por causa do cão. *(Ele) a assustou.

27. A criança está chorando por causa da pintura. *(Esta) a assustou.

CASTILIAN SPANISH

1. Toda responsabilidad queda conmigo. (Yo) tengo que aceptar las consecuencias.
4. Toda responsabilidad queda con Juan S. (Yo) tengo que aceptar las consecuencias.

7. Toda responsabilidad queda con el autor presente. (Yo) tengo que aceptar las consecuencias.

10. Toda responsabilidad queda contigo. (Tú) tienes que aceptar las consecuencias.

16. N/A

19. Toda responsabilidad queda con él. (El) tiene que aceptar las consecuencias.

20. El niño está llorando por causa de este. *(El) lo asustó.

21. El niño está llorando por causa de esta. (Esta) lo asustó.

22. Toda responsabilidad queda con Juan S. (El) tiene que aceptar las consecuencias.

23. El niño está llorando por causa de Rex. *(El) lo asustó.

24. El niño está llorando por causa de la Picasso. (Esta) lo asustó.

25. Toda responsabilidad queda con el chico. (El) tiene que aceptar las consecuencias.

26. El niño está llorando por causa del perro. *(El) lo asustó.

27. El niño está llorando por causa de la pintura. (Esta) lo asustó.


CATALAN

1. Tota responsabilitat està amb mi. (Jo) he de acceptar les consèguencies. 

4. Tota responsabilitat està amb Joan S.. (Jo) he de acceptar les consèguencies.

7. Tota responsabilitat està amb l’autor present. (Jo) he de acceptar les consèguencies.

10. Tota responsabilitat està amb tu. (Tú) has de acceptar les consèguencies.

13. N/A

16. N/A

19. Tota responsabilitat està amb ell. (Ell) ha de acceptar les consèguencies.
20. El bebé està plorant a causa de aquest. (Ell) l'ha espantat.
22. Tota responsabilitat està amb Joan S.. (Ell) ha de acceptar les consèguències.
23. El bebé està plorant a causa de aquest. (Ell) l'ha espantat.
24. El bebé està plorant a causa de la Picasso. (Esta) l'ha espantat.
25. Tota responsabilitat està amb el noi. (Ell) ha de acceptar les consèguències.
26. El bebé està plorant a causa del gos. (Ell) l'ha espantat.
27. El bebé està plorant a causa de la pintura. (Esta) l'ha espantat.

EUROPEAN PORTUGUESE

1. Toda responsabilidade fica comigo. (Eu) tenho de aceitar as conseqüências.
4. Toda responsabilidade fica com o João S. (Eu) tenho de aceitar as conseqüências.
7. Toda responsabilidade fica com o autor presente. (Eu) tenho de aceitar as conseqüências.
10. Toda responsabilidade fica contigo. (Tu) tens de aceitar as conseqüências.
13. Toda responsabilidade fica com o João S. (Tu) tens de aceitar as conseqüências.
16. Toda responsabilidade fica com o médico. (Tu) tens de aceitar as conseqüências.
19. Toda responsabilidade fica com ele. *(Ele) tem de aceitar as conseqüências.
20. A criança está a chorar por causa de aquele. *(Ele) assustou-a.
22. Toda responsabilidade fica com o João S. *(Ele) tem de aceitar as conseqüências.
23. A criança está a chorar por causa do Rex. *(Ele) assustou-a.
25. Toda responsabilidade fica com o menino. *(Ele) tem de aceitar as conseqüências.

ITALIAN

1. Tutta responsabilità sta con me. *(Io) devo accettare le conseguenze.
4. Tutta responsabilità sta con Gianni S.. *(Io) devo accettare le conseguenze.
7. Tutta responsabilità sta con l’autore presente. *(Io) devo accettare le conseguenze.
10. Tutta responsabilità sta con te. (Tu) devi accettare le conseguenze.
13. N/A
16. N/A
19. Tutta responsabilità sta con lui. *(Lui) deve accettare le conseguenze.
20. Il bambino sta piangendo per causa di questo. *(Lui) l’ha spaventato.
21. Il bambino sta piangendo per causa di questa. *(Questa) l’ha spaventato.
22. Tutta responsabilità sta con Gianni S.. *(Lui) deve accettare le conseguenze.
23. Il bambino sta piangendo per causa di Rex. *(Lui) l’ha spaventato.
24. Il bambino sta piangendo per causa della Picasso. *(Questa) l’ha spaventato.
25. Tutta responsabilità sta con il ragazzo. *(Lui) deve accettare le conseguenze.
26. Il bambino sta piangendo per causa del cane. *(Lui) l’ha spaventato.
27. Il bambino sta piangendo per causa della pittura. *(Questa) l’ha spaventato.

PUERTO RICAN SPANISH

1. Toda responsabilidad queda conmigo. *(Yo) tengo que aceptar las consecuencias.
4. Toda responsabilidad queda con Juan S.. *(Yo) tengo que aceptar las consecuencias.
7. Toda responsabilidad queda con el autor presente. (Yo) tengo que aceptar las consecuencias.

10. Toda responsabilidad queda contigo. (Tú) tienes que aceptar las consecuencias.

13. N/A

16. N/A

19. Toda responsabilidad queda con él. *(El) tiene que aceptar las consecuencias.

20. El niño está llorando por causa de este. *(El) lo asustó.

21. El niño está llorando por causa de esta. *(Esta) lo asustó.

22. Toda responsabilidad queda con Juan S.. *(El) tiene que aceptar las consecuencias.

23. El niño está llorando por causa de Rex. *(El) lo asustó.

24. El niño está llorando por causa de la Picasso. *(Esta) lo asustó.

25. Toda responsabilidad queda con el chico. *(El) tiene que aceptar las consecuencias.

26. El niño está llorando por causa del perro. *(El) lo asustó.

27. El niño está llorando por causa de la pintura. *(Esta) lo asustó.
APPENDIX C

BRAZILIAN PORTUGUESE

33. O João leva a Maria para o hospital hoje. *(Ela) está grávida há oito meses.

34. O João leva a Maria para o hospital hoje. *(Ela) está prestes a parir.

35. O João sai com a Maria agora. *(Ela) é muito bonita.

36. O João leva a Maria para o hospital hoje. *(Ela) ainda se sente doente.

37. O João sai com a Maria agora. *(Ela) se estabeleceu recentemente como a primeira bailerina deste país.

38. O João leva a Maria para o centro comercial hoje. *(Ela) vai se comprar novos vestuários de mulher grávida.

39. O João está esperando a Maria há muito tempo agora. *(Ela) está muito ocupada a se fazer bonita para o baile.

40. O João pode levar a Maria para o supermercado. *(Ela) se comporta muito bem em público.

CASTILIAN SPANISH

33. Juan lleva a Maria al hospital hoy. (Ella) está embarazada de ocho meses.

34. Juan lleva a Maria al hospital hoy. (Ella) está a punto de parir.

35. Juan sale con Maria ahora. (Ella) es muy bella.

36. Juan lleva a Maria al hospital hoy. *(Ella) se siente mal todavía.

37. Juan sale con Maria ahora. (Ella) se creó una reputación de la primera bailerina de este país.

38. Juan lleva a Maria al centro comercial hoy. (Ella) va a comprarse nuevos vestidos premamá.

39. Juan está esperando a Maria hace mucho tiempo ahora. (Ella) está ocupada en hacerse bella para el baile.

40. Juan puede llevar a Maria al supermercado. *(Ella) se porta bien en público.
CATALAN

33. En Joan porta la Maria al hospital avui. (Ella) està embarassada desde fa vuit meses.

34. En Joan porta la Maria al hospital avui. (Ella) està a punt de parir.

35. En Joan sort amb la Maria ara. (Ella) és molt bonica.

36. En Joan porta la Maria al hospital avui. (Ella) se sent mal encara.

37. En Joan sort amb la Maria ara. (Ella) s'ha establí com la primera ballerina de aquest país.

38. En Joan porta la Maria al centre comercial avui. (Ella) es compra mous vestits de maternitat.

39. En Joan està esperant la Maria desde fa molt temps ara. (Ella) està ocupada a fer-se bonica per el ball.

40. En Joan pot portar Maria al supermercat. (Ella) es comporta bé en públic.

EUROPEAN PORTUGUESE

33. O João leva a Maria para o hospital hoje. *(Ela) está grávida há oito meses.

34. O João leva a Maria para o hospital hoje. *(Ela) está prestes a parir.

35. O João sai com a Maria agora. *(Ela) é muito bonita.

36. O João leva a Maria para o hospital hoje. *(Ela) ainda sente-se doente.

37. O João sai com a Maria agora. *(Ela) estabeleceu-se recentemente como a primeira bailerina deste país.

38. O João leva a Maria para o centro comercial hoje. *(Ela) vai comprar-se novos vestuários de mulher grávida.

39. O João está a esperar a Maria há muito tempo agora. *(Ela) está muito ocupada a fazer-se bonita para o baile.

40. O João pode levar a Maria para o supermercado. *(Ela) comporta-se muito bem em público.
33. Gianni accompagna Maria all’ospedale oggi. (Lei) è incinta da otto mesi.

34. Gianni accompagna Maria all’ospedale oggi. (Lei) sta per dare all’luce un figlio.

35. Gianni esce con Maria adesso. (Lei) è molto bella.

36. Gianni accompagna Maria all’ospedale oggi. *(Lei) si sente male ancora.

37. Gianni esce con Maria adesso. (Lei) si è stabilita come prima ballerina di questo paese.

38. Gianni accompagna Maria al centro commerciale oggi. (Lei) si compra degli abiti per gestanti.

39. Gianni aspetta Maria da molto tempo adesso. (Lei) è occupata a farsi bella per il ballo.

40. Gianni può accompagnare Maria al supermercato. *(Lei) si comporta bene in pubblico.

---

**PUERTO RICAN SPANISH**

33. Juan lleva a Maria al hospital hoy. *(Ella) está embarazada de ocho meses.

34. Juan lleva a Maria al hospital hoy. *(Ella) está por parir.

35. Juan sale con Maria ahora. *(Ella) es muy bella.

36. Juan lleva a Maria al hospital hoy. *(Ella) se siente mal todavía.

37. Juan sale con Maria ahora. *(Ella) se creó una reputación de la primera bailerina de este país.

38. Juan lleva a Maria al centro comercial hoy. *(Ella) va a comprarse nuevos vestidos premamá.

39. Juan está esperando a Maria hace mucho tiempo ahora. *(Ella) está ocupada en hacerse bella para el baile.

40. Juan puede llevar a Maria al supermercado. *(Ella) se porta bien en público.
REFERENCES


---


262


---


