

ANAPHORS AT THE INTERFACES.
A COMPARATIVE STUDY OF THE VARIATION OF
THE ANAPHORIC SYSTEMS OF ENGLISH, DUTCH AND SPANISH

Ismael Iván Teomiro García

Doctoral Dissertation



Advisors:

Dr. Amaya Mendikoetxea Pelayo

Dr. Olga Fernández Soriano



Programa de doctorado "Lingüística Teórica y Sus Aplicaciones"

Instituto Universitario de Investigación Ortega y Gasset



Departamento de Lingüística, Lenguas Modernas, Lógica y Filosofía de

la Ciencia y Teoría de la Literatura y Literatura Comparada

Facultad de Filosofía y Letras

Universidad Autónoma de Madrid

Madrid 2010

Acknowledgments

I am greatly indebted to my advisors, Olga Fernández Soriano and Amaya Mendikoetxea for their constant help and support. They have led me along the way, and pointed out the goal when it was right before me but I could not see it. I thank Amaya also for the opportunity to work for the Department of English Studies of the Universidad Autónoma de Madrid. I will ever always be grateful for that.

I would also like to thank the members of the WOSLAC project: Cristóbal Lozano, Susana Murcia, Mick O'Donnell, Rocío Jiménez, Esther Ferrandis, Paul Rollinson, and, especially, Gema Chocano, who has helped me improve some aspects of the present dissertation. I am also grateful to the other members of the SINTECOM group (F-064) at the Universidad Autónoma de Madrid: Luis Eguren and especially Silvia Serrano Pardo for her advice on stylistic issues. Lourdes Muñoz has helped me with a lot of administrative matters. I am also indebted to the members of my T.E.A. (Tribunal de Estudios Avanzados) committee: Ricardo Mairal and Carlos Piera. I gratefully acknowledge the financial support of the Ministry of Science and Innovation of Spain (Ministerio de Ciencia e Innovación de España): grant BES-2006-11703, project "Las Interficies Léxico-Sintaxis y Sintaxis-Discurso: Factores Sintácticos y Pragmáticos en la Adquisición del Orden de Palabras en Inglés y Español como Segundas Lenguas" (HUM2005-01728), and project "Opcionalidad y Pseudo-opcionalidad en las Gramáticas Nativas y No Nativas" (FFI2008-01584/FILO).

I am also very grateful to other people whose help has been very important for the development of this thesis: Silvia Serrano Pardo, Cristina Romero, and Violeta Martínez (I.U.I Ortega y Gasset); Violeta Demonte, Isabel Pérez Jiménez and Héctor Alcalde (CSH-CSIC); Jose María Brucart, Gemma Rigau, Carme Picallo, Maria Lluïsa Hernanz, Jaume Mateu, Maria Teresa Espinal, Elías Gallardo, Yurena Guitérrez, Verónica Castillo-Herrerías and Silvia Ferreiro (CLT-UAB); Eric Reuland, Martin Everaert, Marijana Marelj and Rianne Schippers (UiL-OTS, Utrecht University); Ana Ardid (UAM), Christina Sevdali, Johanna Sattler and Tania Strahan.

Table of Contents

1. Introduction	1
2. Data and hypotheses	5
2.1. Data.....	5
2.2. Hypotheses.....	15
2.3. Theoretical framework.....	17
3. Control and null SE-anaphors at the interfaces	23
3.1. Introduction.....	23
3.2. Current approaches to control.....	25
3.2.1. The Null Case Theory	26
3.2.2. The Movement Theory of Control	29
3.2.2.1. Hornstein (1999)	29
3.2.2.2. Manzini & Roussou (2000).....	31
3.2.2.3. Theoretical problems.....	33
3.2.2.4. Empirical problems	34
3.2.3. The Calculus of Control	35
3.2.4. Recapitulation.....	39
3.3. Theoretical background	39
3.3.1. Anaphors and anaphora resolution	39
3.3.1.1. Introduction: Deixis and anaphora	40
3.3.1.2. Strategies of anaphora resolution	41
3.3.1.3. Binding Theory	44
3.3.1.4. Summary	49
3.3.2. Anaphors and binding domains.....	49
3.3.2.1. Beyond the Binding Theory: Control and PRO	51
3.3.2.2. Locality and condition B: Disjoint Reference Effect and short distance pronouns.....	53
3.3.2.3. Locality and condition A: Long distance anaphora	56
3.3.2.4. Summary	60
3.3.3. Anaphors, binding and syntactic chains	60
3.3.4. Agree-chains.....	63
3.4. The null SE-anaphor PRO	65
3.4.1. The interpretation of PRO	65
3.4.1.1. Obligatory control	66
3.4.1.2. Non obligatory control	69
3.4.1.3. Some notes on arbitrary control	70
3.4.2. The distribution of PRO	71
3.4.2.1. Case and silent PRO	71
3.4.2.2. English infinitive clauses	73
3.4.2.3. Spanish infinitive clauses	74
3.5. English <i>ing</i> -clauses	78
3.5.1. Pires' (2001) analysis.....	79
3.5.2. The alternative analysis	82
3.6. Spanish infinitives: NcI vs. PRO	88

3.6.1. P+infinitive.....	89
3.6.2. Infinitive in subject position.....	91
3.6.3. Exclamative and interrogative infinitives.....	94
3.7. Notes on some other issues.....	95
3.7.1. Little <i>pro</i>	95
3.7.2. OC in Hebrew finite clauses.....	97
3.7.3. European Portuguese inflected infinitives.....	98
3.7.4. OC into adjuncts.....	99
3.8. Conclusions.....	100
4. Romance <i>se/si</i> clitics.....	103
4.1. Introduction.....	103
4.1.1. Clitics and cliticization in Romance.....	103
4.1.2. Romance <i>se/si</i> clitics.....	106
4.1.3. Constructions with <i>se</i> in Spanish.....	110
4.2. Impersonal and passive SE.....	113
4.2.1. Literature review.....	113
4.2.2. Recapitulation and hypotheses.....	125
4.3. Middle SE.....	126
4.3.1. Literature review.....	126
4.3.2. Recapitulation.....	128
4.4. Reflexive SE.....	128
4.4.1. Literature review.....	129
4.4.2. Recapitulation and hypotheses.....	132
4.5. Inchoative / ergative SE.....	132
4.5.1. Literature review.....	133
4.5.2. Recapitulation and hypotheses.....	135
4.6. Inherent SE.....	135
4.6.1. Literature review.....	135
4.6.2. Recapitulation and hypotheses.....	140
4.7. Recapitulation and implications.....	140
4.7.1. Recapitulation.....	140
4.7.2. <i>Se</i> in Slavic.....	144
4.7.3. Romance <i>se/si</i> and Germanic languages.....	147
4.7.4. From null to overt SE-anaphors.....	147
5. Reflexivity and overt SE-anaphors at the interfaces.....	149
5.1. Introduction: Reflexivity in Romance and Germanic.....	149
5.1.1. Reflexivity in English and Dutch.....	149
5.1.2. Reflexivity in Spanish.....	152
5.1.3. Recapitulation.....	160
5.2. Theoretical background: The lexicon-syntax interface and reflexivization.....	160
5.2.1. Beyond the Canonical Binding Theory: SE- and SELF-anaphors.....	161
5.2.2. Active lexicon and the Theta System.....	165
5.2.2.1. Lexical entries, theta-features and theta-clusters.....	166
5.2.2.2. Lexical marking and syntactic merging.....	167
5.2.2.3. Theta-chains.....	168
5.2.2.4. Active lexicon and arity operations I: Reflexivization.....	169

5.2.3. Null SE-anaphors	173
5.2.4. Recapitulation: Lexical reflexivization and reflexive binding	174
5.3. SE-anaphors at the lexicon-syntax and syntax-semantics interfaces	177
5.3.1. Inherent reflexive verbs, SE-anaphors and the Double Chain Condition ..	178
5.3.2. Non inherent reflexive verbs and SELF-anaphors	184
5.3.3. Oblique <i>sí</i>	187
5.3.4. SELF-anaphors vs. SE-anaphors with Inherent Reflexive Verbs.....	189
5.3.5. SELF-logophors.....	191
5.3.5. Recapitulation.....	192
5.4. SE-anaphors at the syntax-phonology interface.....	192
5.5. Conclusions.....	194
6. Pronominal verbs, variation and adjustment strategies at the interfaces	197
6.1. Introduction.....	197
6.1.1. Preliminaries.....	198
6.1.2. Data	199
6.1.3. Theoretical hypotheses	203
6.1.4. Questions to answer.....	204
6.1.5. Working hypotheses	204
6.2. Theoretical background: The lexicon-syntax interface and unaccusativity.....	205
6.3. Pronominal verbs	207
6.3.1. Inherent reflexive verbs.....	207
6.3.2. Decausativized experiencer verbs	211
6.3.3. Decausativized theme verbs	214
6.3.4. Cross-linguistic differences at the syntax-phonology interface	217
6.4. Non-pronominal verbs	220
6.4.1. Internally caused unaccusative verbs: <i>floreecer</i> -type verbs.....	221
6.4.2. Internally caused unaccusative verbs: <i>augmentar</i> -type verbs.....	223
6.4.3. Existential and presentational unaccusative verbs	225
6.4.4. Agentive unergative verbs.....	227
6.4.5. Consumption verbs.....	230
6.4.6. Double caused verbs: <i>cambiar</i>	233
6.5. Alternating pronominal verbs	233
6.5.1. Verbs with animacy restrictions: [+c][-c] verbs.....	235
6.5.1.1. <i>Mejorar(se)</i> / (<i>zich</i>) <i>bewegen</i> -type verbs	235
6.5.1.2. <i>Envejecer(se)</i> -type verbs	237
6.5.2. Verbs without animacy restrictions: Theme-subject matter verbs	238
6.6. Aspectual <i>se</i>	242
6.7. Complex [+] entries	245
6.8. Conclusions.....	246
7. Recapitulation and implications: SE-anaphors in arbitrary constructions....	251
7.1. Recapitulation	251
7.2. Arbitrary constructions	253
7.2.1. Arbitrary control.....	253
7.2.2. Impersonal and passive SE.....	254
7.2.3. The semantics of arbitrary: Review.....	255

7.3. SE-anaphors at the syntax-semantics interface	265
7.3.1. Nominative vs. dative SE-anaphors	266
7.3.1.1. ARB PRO is null PRO	266
7.3.1.2. Impersonal SE is overt nominative PRO'	267
7.3.1.3. Passive SE is overt dative PRO'	268
7.3.1.4. Arbitrary SE in Germanic	269
7.3.2. Arbitrary interpretation and choice functions	269
7.4. Conclusions.....	272
8. Conclusions.....	273
References.....	281
Abstract.....	293
Resumen.....	297
Samevatting	301

Chapter 1

Introduction

Human languages make use of nominal elements in order to denote entities in the world. These nominal items can refer either to entities that form part of the communicative situation (deixis) or to entities that do not form part thereof (reference). Any theory of language has to account for the interpretation of nominal elements, whether it is deictic or referential.

In Generative Grammar, particularly in the framework of Government & Binding, Chomsky (1981) proposed the Binding Theory in order to account for the referential properties of nominal items. This theory was one of the components that made up the modular grammar pursued in the aforementioned framework. Later on, Reinhart & Reuland's (1993) version of the Binding Theory could also account for some deictic interpretations of certain nominal elements.

The Binding Theory proposed by Chomsky (1981) is specified in (1). It was proposed in order to account for the referential properties of anaphoric elements like those in sentences (2) and (3). The element *himself* in (2) (an *anaphor*, in Chomsky's terms) has to refer to *John*, and no other antecedent is possible (i.e. it cannot refer to Bill). In sentence (3), the referential properties of the pronoun *him* (a *pronominal* in Chomsky's terms) is completely different: it cannot refer to *John* but its antecedent has to be anyone except for *John* (e.g. Bill).

- (1) Binding Theory: (Chomsky 1981:188)
- A. An anaphor is bound in its governing category.
 - B. A pronominal is free in its governing category.
 - C. An R-expression is always free.
- (2) Bill said that John washed himself.
- (3) Bill said that John washed him.

Although the original Binding Theory covered much of the data scholars were aware of at the time, it soon run into some crucial problems that nowadays have not yet been solved in a satisfactory way.

First, the existence of the null pronoun PRO as in (4) has always been an important difficulty for the Binding Theory.

(4) Bill wanted PRO to read the book.

In order to account for its syntactic distribution, PRO had to be defined both as an anaphor and as a pronominal at the same time. It is in this sense that Chomsky (1982) defined PRO as a *pronominal anaphor*, whose syntactic behaviour could not be accounted for by the standard Binding Theory. Therefore, Control Theory was proposed as an independent module of the grammar in order to explain the nature, syntactic behaviour and referential interpretation of PRO.

Another problem was the fact that there are languages that make use of other anaphoric elements that could not be accounted for by the standard Binding Theory either. For example, Dutch has a three-way anaphoric system (Everaert 1986) rather than a two-way anaphoric system as proposed by Chomsky (1981) in (1) above. Besides *anaphors* and *pronominals*, Dutch also has *SE-anaphors*, whose syntactic distribution does not obey Chomsky's (1981) Condition A or Condition B. Spanish, as well as other Romance languages, makes use of anaphoric elements that not only appear in reflexive environments but also in other configurations where they do not seem to have an anaphoric interpretation, such as in ergative constructions (5) and impersonal constructions (6) (see Belletti 1982; Cinque 1988; Otero 1986, 1999; Mendikoetxea 1992, 1999; among many others).

(5) El vaso **se** rompió.

The glass *se* broke

"The glass broke."

(6) **Se** dice que alguien lo hizo a propósito.

se says that someone *it_{clitic}* did on purpose

"People say someone did it on purpose."

The starting point of this thesis is the Control Theory and control constructions like (4). I will then move on to study other constructions such as reflexives (2), ergatives (5) and impersonals (6). In doing so, the model will be able to account for the nature, syntactic behavior and semantic interpretation of all the anaphoric elements present in Romance (Spanish) and Germanic languages (English and Dutch) in a unified way along the lines of Chomsky's (1995, 2001, 2005) Minimalist Program. The ultimate goal of this work is

to give a unified account of a series of constructions that previously had been studied as independent from one another, such as impersonals, reflexives, and ergatives.

The thesis is structured as follows. In chapter 2 I will present the data that motivate the thesis, the hypotheses that will be put forward and the theoretical background upon which the analysis to be developed will be based.

In chapter 3 I will centre on null SE-anaphors. I will address the issue of control phenomena and the pronominal anaphor PRO. My claim is that PRO is a SE-anaphor without phonological content (i.e. it is a null element) and with its φ -features unvalued. Depending on where the binding of PRO takes place (syntax or semantics), obligatory or non-obligatory control result (respectively). I will also revolve around the syntactic behaviour and semantic interpretation of arbitrary control and PRO_{arb}, as well as impersonal and passive constructions with SE/SI/SIĒ. Nevertheless, these latter issues will not be fully addressed until chapter 7.

In chapter 4 I will review the most relevant literature on Romance and Slavic clitics in Generative Grammar. I will pay attention mainly to Italian and Spanish. Furthermore I will briefly discuss other Romance languages, such as French, European Portuguese and Romanian, as well as the Slavic family.

In chapter 5 I will concentrate on overt SE-anaphors by addressing reflexivization in general, and more concretely, reflexivization in Spanish. I will introduce the null SE-anaphor PRO' and its different phonological realisations (*se/si/siĒ*). The main difference of PRO' compared to PRO is that the former has no grammatical number feature, and hence, no possibility of being referentially independent. I will show that reflexivization in Spanish follows the same rules as in other languages like English and Dutch. Those rules will be formalized in Reinhart and Reuland's (1993) Conditions A and B.

In chapter 6 I will focus on pronominal verbs. These are reflexive and unaccusative verbs whose argument structure requires the presence of a reflexive clitic that does not seem to be interpreted as an anaphor at the semantic level. I will demonstrate that all these verbs undergo different operations at the lexicon-syntax interface. Moreover, I will show that the agree-clitic is a defective SE-anaphor (namely PRO') inserted for the convergence of the derivation at the linguistic interfaces rather than for interpretive reasons. In conclusion, I will account for the variation of the realization of pronominal verbs by resorting to processes and adjustment strategies at the interfaces.

In chapter 7 I will address the issue of the role of anaphors in arbitrary constructions. Arbitrary control, impersonal *se/si* and passive *se/si* will be analysed by means of the presence of a SE-anaphor in subject position, which is interpreted by means of a choice function introduced by the anaphor and this results in the arbitrary interpretation.

Finally, I will present the conclusions of the dissertation in chapter 8.

Chapter 2

Data and hypotheses

In this chapter I will first present the empirical data that the analysis to be developed in this dissertation will account for. Subsequently, I will propose the hypotheses that will lead the research in the following chapters. Thereafter, I will introduce the theoretical framework and the model of the language upon which my analysis will be built.

2.1. Data

In this section I will review several constructions (control phenomena, reflexivity and reciprocity, passives and impersonals, among others) in different languages that belong to Germanic, Romance and Slavic families. I will briefly describe the constructions and their main syntactic properties.

First, I will take a look at control phenomena, which have been the subject of investigation in Generative Grammar since its earliest days (Rosenbaum's (1967) *Equi-NP Deletion*). In the framework of Government & Binding (Chomsky 1981) the analysis of control phenomena was based on the null element PRO, which was analyzed as a pronominal anaphor in subject position. Typical examples of control phenomena are given below. Sentences in (1) show what is known as *obligatory control* (OC): (1a) in English, (1b) in Dutch, and (1c) in Spanish. In OC environments the null subject of the embedded sentence has to be co-referential with the subject of the matrix clause.

- (1) a. I_i want [PRO_{i/*j/*arb} to read the Silmarillion].
b. Ik_i wil [PRO_{i/*j/*arb} de Silmarillion lezen].
I want [PRO the Silmarillion read]
"I want to read the Silmarillion."
c. Yo_i quiero [PRO_{i/*j/*arb} leer el Silmarillion].
I want [PRO read the Silmarillion].
"I want to read the Silmarillion."

On the other hand, sentences in (2) are examples of *non-obligatory control* (NOC): (2a) in English, (2b) in Dutch and (2c) in Spanish. In NOC environments the embedded subject can be correferential with the matrix subject, although it does not have to.

- (2) a. John_i's friend_j thinks that [**PRO**_{i/j/k/arb} being naive is something bad].
 b. Jan_is vriend_j denkt dat [het **PRO**_{i/j/k/arb} naïef zijn iets ergs is].
 John's friend thinks that [the **PRO** naive be something bad is].
"John's friend thinks that being naive is something bad."
 c. El amigo_j de Juan_i piensa que [**PRO**_{i/j/k/arb} ser ingenuo es malo].
 The friend of Juan thinks that [**PRO** be naive is bad].
"John's friend thinks that being naive is something bad."

Sentences that have a null PRO subject, and that receive an arbitrary interpretation are the result of what has been called *arbitrary control* (AC)¹. The verb, as in the cases of OC and NOC, appears in its infinitival form (either *to*-infinitive or *ing*-infinitive):

- (3) a. [**PRO**_{arb} reading the Silmarillion] is difficult.
 b. [Het **PRO**_{arb} lezen van de Silmarillion] is moeilijk.
 [The **PRO** read of the Silmarillion] is difficult.
"Reading the Silmarillion is difficult."
 c. [**PRO**_{arb} Leer el Silmarillion] es difícil.
 [**PRO** Read the Silmarillion] is difficult.
"Reading the Silmarillion is difficult."

Not only by means of AC can arbitrary subjects be expressed in Romance and some Slavic languages. They can also make use of an **agree-clitic** *si*²/*se*³/*się*⁴ in subject position rather than PRO_{arb}. As in the cases of AC, the clitic may receive either a quasi-universal (or generic) interpretation or a quasi-existential interpretation, depending on the presence of certain operators in the sentence⁵. This construction is called *impersonal se/si/się* when the verb shows up with its default third person agreement, i.e. there is no

¹ See Manzini & Roussou (2000:427).

² For Italian, see Belletti (1982), Burzio (1980, 1986), and Cinque (1988) among many others.

³ For Spanish, see Otero (1986) and Mendikoetxea (1992, 1999, 2008), among many others. For French see Kayne (1975, 1991, 2000). For Portuguese, see Raposo & Uriagereka (1996). For Romanian, see Dobrovie-Sorin (1998, 2006).

⁴ For Slavic languages in general, and Polish in particular, see Rivero (2001).

⁵ For further discussion, see Cinque (1988), Chierchia (1995), and Mendikoetxea (2002, 2008).

verb-object agreement. Sentences (4a,b) show impersonal *se* in Spanish, impersonal *si* in Italian can be seen in (4c,d), sentence (4g) shows impersonal *se* in Portuguese, and (4j,k) illustrates impersonal *się* in Polish. Note that this particular construction (without verb-object agreement) is not possible in French (4e,f) or Romanian (4h,i)⁶.

- (4) a. Aquí **se** lee libros. (Mendikoetxea 2008:316)
 Here **se** read_{singular} books
"One (SE) reads books (here)."
- b. **Se** baila mucho en las fiestas. (Mendikoetxea 2008:303)
Se dance_{singular} a lot in the parties
"One (SE) dances a lot at parties."
- c. Gli **si** telefona spesso. (Burzio 1986:43)
 Him_{dative} **si** phones often
"We phone him often."
- d. **Si** leggerà volentieri alcuni articoli. (Burzio 1986:43)
Si will read_{sing} willingly a few articles
"We will be eager to read a few articles."
- e. ***Se** mange les noix. (Belletti 1982:20)
 ***Se** eat the hazelnuts
"One eats hazelnuts."
- f. ***Se** va à Paris. (Belletti 1982:20)
 ***Se** go to Paris
"One goes to Paris."
- g. Compra-**se** sempre demasiadas salsichas no talho Sanzot. (R&U 1996:750⁷)
Se-buy_{sing} always too-many sausages at the butcher shop Sanzot
"One (people) always buy too many sausages at the Sanzot butcher shop."
- h. *Nu **se** este niciodată mulțumit. (Dobrovie-Sorin 1998:405)
 Not **se** is never satisfied
"One is never satisfied."

⁶ See Belletti (1982), Cinque (1988), Mendikoetxea (1992) and Dobrovie-Sorin (1998, 2006) for more detailed analyses and further discussion.

⁷ R&U (1996) stands hereinafter for Raposo & Uriagereka (1996).

- i. *In această universitate **se** predă științele umane. (Dobrovie-Sorin 1998:405)
 In this university **se** teaches the humanities
"In this university one teaches the humanities."
- j. Teț książkę (czyta / czyta**o**) **się** z przyjemnością. (Rivero 2001:171)
 This book_{acc} (read_{3rd,sing} / read_{neu}) **się** with pleasure
"One (reads / read) this book with pleasure."
- k. Tutaj **się** pracuje sporo. (Rivero 2001:170)
 Here **się** works much
"Here people work a lot."

These constructions can appear with agreement between the verb and the object (so the verb must be transitive), rather than the default 3rd person singular features (Spanish). Sentences in (5a,b) show this construction in Spanish, (5c,d) in Italian and (5g,h) in Portuguese. These cases, which are called *passive se/si/się*, can also be attested in French (5e,f) and Romanian unlike the impersonal constructions in (4e,f).

- (5) a. **Se** pasaron los trabajos a ordenador. (Mendikoetxea 1999:1635)⁸
Se wrote the papers to computer
"The papers were typed in the computer."
- b. **Se** comen las manzanas. (Mendikoetxea 2008:291)
Se eat_{plural} the apples
- c. **Si** mangiano le mele. (Mendikoetxea 2008:291)
Si eat_{plural} the apples
- d. **Si** leggeranno volentieri alcuni articoli. (Burzio 1986:43)
Si will read_{plural} willingly a few articles
"A few articles will be read eagerly."
- e. Il_j s'**est** traduit trois romans_i. (Dobrovie-Sorin 2006:122)
 it **se**-has translated three novels
"Three novels were translated."
- f. Les noixettes **se** mangent. (Belletti 1982:19)
 The hazelnuts **se** eat_{plural}
"The hazelnuts are eaten."

⁸ The glosses and translation are mine.

- g. Ontem compraram-se demasiadas salsichas no talho Sanzot. (R&U 1996:750)
 Yesterday **se**-bought_{plur} too many sausages at the butcher-shop Sanzot
"Yesterday someone or other bought too many sasages at the Sanzot butcher shop."
- h. Copiaram-se antes de destruímos ec. (R&U 1996:790)
Se-copied_{plural} before of to-destroy_{1st.plural}
"Someone or other copied them before we destroyed them"

Also in middle constructions⁹ this agree-clitic appears, as can be seen in (6a) for Spanish, (6b) for Italian, (6c) for French, and (6d) for Polish.

- (6) a. Este libro **se** lee facilmente.
 This book **se** reads easily
"This book reads easily."
- b. Questo tavolino **si** transporta facilmente. (Cinque 1988:559)
 This table **si** transports easily
"This table transports easily (is easily to transport)"
- c. Le grec **se** traduit facilement. (Dobrovie-Sorin 2006:122)
 The Greek **se** translates easily
"Greek translates easily."
- d. Ten samochòd powadzi **się** łatwo. (Rivero 2001:170)
 This car drives **się** easily
"This car drives easily."

The cases of *se/si/się* described just above are what Burzio (1986) referred to SI, which is never inflected. Nonetheless, this clitic appears in other constructions too, where it must be inflected for person and number. This is what Burzio referred to *si*, and it appears with reflexive (7), reciprocal (8), inchoative or ergative (9), and inherent reflexive (10) verbs in several languages.

The sentences in (7a,b) show reflexive *si* in Italian. Reflexive *se* in Spanish can be seen in (7c). Portuguese reflexive *se* appears in (7d), and finally (7e) and (7f) show reflexive *se* and *się* for French and Romanian respectively.

⁹ For a definition of *middle construction* as will be understood in this thesis, see chapter 4 section 4.3.

- (7) a. Maria **si** guarda. (Burzio 1986:37)
Maria **si** watches
"Maria watches herself"
- b. Sarebbe bello [PRO vedersi più spesso]. (Burzio 1986:51)
Would be nice [PRO see **si** more often].
"It would be nice to see each other more often."
- c. Juan **se** lavó.
Juan **se** washed
"Juan washed (himself)."
- d. Os meninos insultaram-**se**. (R&U 1996:801)
The children insulted-**se**
"The children insulted themselves (each other)."
- e. Jean **se** lave. (Dobrovie-Sorin 2006:123)
Jean **se** washes
"Jean washes himself."
- f. Janek ubiera **się**. (Rivero 2001:170)
John dresses **się**
"John gets dressed."

The sentences in (8a,b) show reciprocal *se* in Spanish, while reciprocal *se* in Portuguese can be seen in (8c).

- (8) a. Los padres **se** despidieron
The parents **se** said goodbye
"The parents said goodbye to each other."
- b. Los niños *(**se**) lavaron (el uno al otro).
The children **se** washed (the one to the other)
"The children washed (one another)."
- c. Os meninos insultaram-**se**. (R&U 1996:801)
The children insulted-**se**
"The children insulted themselves (each other)."

Inchoative or ergative *se* can be seen in (9a) for Spanish. In (9b) it can be seen for Italian. French ergative *se* is in (9c), and finally (9d) shows ergative *się* in Polish.

- (9) a. El cristal **se** rompió. (Mendikoetxea 2008:291)
 The glass **se** broke
"The glass broke."
- b. Il vetro **si** rompe. (Burzio 1986:37)
 The glass **si** breaks
"The glass breaks (itself)"
- c. La branche s'est cassée. (Dobrovie-Sorin 2006:121)
 The branch **se**-has broken
"The branch broke."
- d. Szklanka **się** rozbila. (Rivero 2001:170)
 Glass **się** broke
"The glass broke."

Finally, inherent *se*¹⁰ can be seen in (10a,b) for Spanish, (10c) for Italian, (10d) for French, and (10e) for Polish.

- (10) a. María **se** asusta de Juan.
 Mary **se** gets scared of John
"Mary gets scared of John."
- b. Juan **se** arrepintió.
 Juan **se** changed his mind
"Juan changed his mind."
- c. Giovanni **si** sbaglia. (Burzio 1986:37)
 Giovanni **si** mistakes
"Giovanni mistakes (himself)"
- d. Marie s'est souvenu de Jean. (Dobrovie-Sorin 2006:169; endnote 10)
 Marie **se**-has remembered of Jean
"Marie remembered Jean."
- e. Maria boi **się** Janka. (Rivero 2001:170)
 Mary fears **się** John
"Mary gets afraid of John."

To sum up so far, we have seen control phenomena and constructions with agree-clitics. As said before, Burzio (1986) distinguished between SI, which is never inflected and

¹⁰ For a definition of *inherent se* see chapter 4 section 4.6.

appears in impersonal, passive and middle constructions, and *si*, which is inflected for person and number, and appears with reflexive, reciprocal, ergative and inherent reflexive verbs. We will follow Burzio's nomenclature and differentiate between SE/SI/SIĒ (for impersonals, passives and middles) and *se/si/siĕ* (for reflexives, reciprocals, ergatives and inherent reflexives).

Moreover, *se* can also appear in certain Spanish constructions, where it seems to have an aspectual meaning, as in (11).

- (11) Ana (**se**) comió las manzanas (Otero 1986:87)
 Ana (**se**) ate the apple
"Ana ate (up) the apples."

Finally, a special instance of clitics in Spanish are *dative clitics*. In these cases, such as (12) below, the agree-clitic behaves as a participant in the event that denotes the experiencer or benefactive of the result of the event. However, this extra participant cannot be considered a verbal argument since it is not selected by the verb. The clitic can never be *se* but can be *me* (1st person singular), *te* (2nd person singular), *le* (3rd person singular), *nos* (1st person plural), or *os* (2nd person plural).

- (12) Este niño no **me** come nada
 This child not **me**_{dative} eats nothing
"This child eats nothing (and it affects me)."

As for Germanic languages, Dutch and German make use of certain kind of anaphors in contexts where Romance uses agree-clitics. For example, in Dutch the anaphor *zich* appears with reflexive (13a), reciprocal (13b), and inherent reflexive verbs (13c). However and unlike Romance, *zich* does not appear with ergative verbs (14a). Nor does this language have a *zich*-counterpart to the Romance SE/SI, i.e. *zich* never appears in middle constructions (14b), and nor do counterparts of Romance impersonal and passive constructions exist in Dutch. Finally, *zich* seems not to affect the aspect of the sentence as seen in (14c).

- (13) a. Jan scheert **zich** elke morgen.
 Jan shaves **zich** every morning
"Jan shaves every morning."

- b. Jan en Marie wassen **zich** elke morgen.
 Jan and Marie wash **zich** every morning
"Jan and Marie wash (each another) every morning."
- c. Jan bedacht **zich**.
 Jan changed his mind **zich**
"Jan changed his meaning."

- (14) a. Het glas is (***zich**) gebroken.
 The glass has **zich** broken
"The glass broke."
- b. Deze schoenen lopen (***zich**) prima.
 These shoes walk **zich** well
"One walks well with these shoes."
- c. Jan heeft (***zich**) de apple gegeten (=Jan heeft de apple **opgegeten**)
 Jan has **zich** the apple eaten (=Jan has the apple **up** eaten)
"Jan has eaten up the apple."

On the other hand, English has neither agree-clitics nor (overt) anaphors for none of the aforementioned constructions or verb types. Sentence (15a) shows a reflexive construction, whereas a reciprocal construction can be seen in (15b). An ergative verb is in (15c), and an inherent reflexive verb is shown in (15d). An example of middle construction is in (15e), and finally we see that the function of the aspectual *se* in Spanish (11) can be fulfilled by the particle *up* in English (15f).

- (15) a. John washed \emptyset .
 b. John and Mary kissed \emptyset .
 c. The glass broke \emptyset .
 d. John worried \emptyset about Max.
 e. The Silmarillion reads \emptyset easily.
 f. John ate **up** the apple.

To conclude this section, I have reviewed some apparently unrelated constructions in some languages. First, I have shown control phenomena, both OC and NOC. Subsequently, I have looked at generic subjects in two different syntactic environments: in AC where the subject is PRO_{arb}, and in impersonal and passive constructions, where the subject is an agree-clitic that is never inflected for person or number. The latter is

attested in Romance languages and some Slavic languages. We have also seen that this agree-clitic is present in middle constructions in the aforementioned languages. Thereafter, I have looked at other constructions in Romance and Slavic that require this agree-clitic though inflected for person and number: reflexive, reciprocal, ergative and inherent reflexive verbs. Another use of the agree-clitic *se* in Spanish has been reviewed in this section too: when *se* has an aspectual meaning.

Finally, we have seen that there are important cross-linguistic differences when looking at Romance and Germanic languages. Dutch, on the one hand, does make use of the anaphor *zich* in some constructions where Romance languages use the agree-clitics: reflexive, reciprocal and inherent reflexive verbs. However, Dutch bans the occurrence of *zich* with ergative verbs, unlike Romance. On the other hand, English seems to use nothing in all the so far reviewed constructions, in sharp contrast with both Romance and Dutch.

Table (16) summarizes the data reviewed so far. The columns represent the different languages or families, whereas the rows represent the different constructions that require the elements referred to in the individual cells.

(16) Summary of the data:

	English	Dutch	Romance	Slavic¹¹
OC	PRO	PRO	PRO	PRO
NOC	PRO	PRO	PRO	PRO
AC	PRO _{arb}	PRO _{arb}	PRO _{arb}	PRO _{arb}
Impersonal	-	-	SE/SI	<i>SIE</i>
Passive	-	-	SE/SI	<i>SIE</i>
Middle	-	-	SE/SI	<i>SIE</i>
Reflexive	-	<i>zich</i>	<i>se/si</i>	<i>się</i>
Reciprocal	-	<i>zich</i>	<i>se/si</i>	<i>się</i>
Ergative	-	-	<i>se/si</i>	<i>się</i>
Inherent reflexive	-	<i>zich</i>	<i>se/si</i>	<i>się</i>
Aspectual¹²	verbal particles	verbal particles	<i>se/si</i>	?
Ethical¹³	?	?	<i>me/te/le/nos/os/les</i>	?

2.2. Hypotheses

A unified analysis of all the constructions in (16) is possible. This is precisely what I will do in this work by putting forward three hypotheses:

- A. All the structures seen above and summarized in (16) have in common the presence of an anaphor (more concretely, a SE-anaphor, as will be argued later on). On the one hand, this anaphor can be null, as in the cases of PRO in OC and NOC, PRO_{arb} in AC, and in some cases in English and Dutch. On the other hand, the anaphor is overt, as in the cases of SE/SI/*SIE* and *se/si/się* in Romance and Slavic languages, and *zich* in Dutch. The table (16) is then redefined in (17) below:

¹¹ See chapter 4 section 4.7.2. and Rivero (2001).

¹² The data analysed here refer just to Spanish.

¹³ The data analysed here refer just to Spanish.

(17) Summary of the data (revisited):

	English	Dutch	Romance	Slavic
OC	PRO	PRO	PRO	PRO
NOC	PRO	PRO	PRO	PRO
AC	PRO _{arb}	PRO _{arb}	PRO _{arb}	PRO _{arb}
Impersonal	does not exist	does not exist	SE/SI	<i>SIE</i>
Passive	does not exist	does not exist	SE/SI	<i>SIE</i>
Middle	null anaphor	null anaphor	SE/SI	<i>SIE</i>
Reflexive	null anaphor	<i>zich</i>	<i>se/si</i>	<i>się</i>
Reciprocal	null anaphor	<i>zich</i>	<i>se/si</i>	<i>się</i>
Ergative	null anaphor	null anaphor	<i>se/si</i>	<i>się</i>
Inherent reflexive	null anaphor	<i>zich</i>	<i>se/si</i>	<i>się</i>
Aspectual ¹⁴	verbal particles	verbal particles	<i>se/si</i>	?
Ethical ¹⁵	?	?	<i>me/te/le/nos/os/les</i>	?

B. The system of concepts¹⁶, the narrow syntax and the semantic system are universal (Chomsky 2001, 2005, 2006; Reinhart 2002; Reinhart & Siloni 2005). So, whether the anaphor is overt or null is something that depends on the interface between the syntax and the Sensory-Motor System (S-M). In the same way, whether the anaphor is interpreted or not as a participant of the event depends on processes at the lexicon-syntax and syntax-semantics interfaces.

C. The cross-linguistic, diachronic and intra-linguistic variation observed among these constructions must be located at the interfaces if we are to maintain the syntax as a universal component (following Chomsky 1995, 2001, 2005), as well as the Conceptual Intentional system (C-I) and the lexicon (Reinhart 2002; Reinhart & Siloni 2005). More concretely I will demonstrate that this variation is due to:

- a. arity operations at the lexicon-syntax interface,
- b. binding processes at the syntax-semantics-discourse interfaces,

¹⁴ The data analysed here refer just to Spanish.

¹⁵ The data analysed here refer just to Spanish.

¹⁶ See section 2.3 below and particularly Reinhart (2002) for a full description of the System of Concepts.

- c. materialization rules at the syntax-phonology interface (i.e. whether the anaphor is overt or null), and
- d. conditions at the syntax-semantics interface (i.e. whether the anaphor is interpreted as a participant of the event or not).

2.3. Theoretical framework

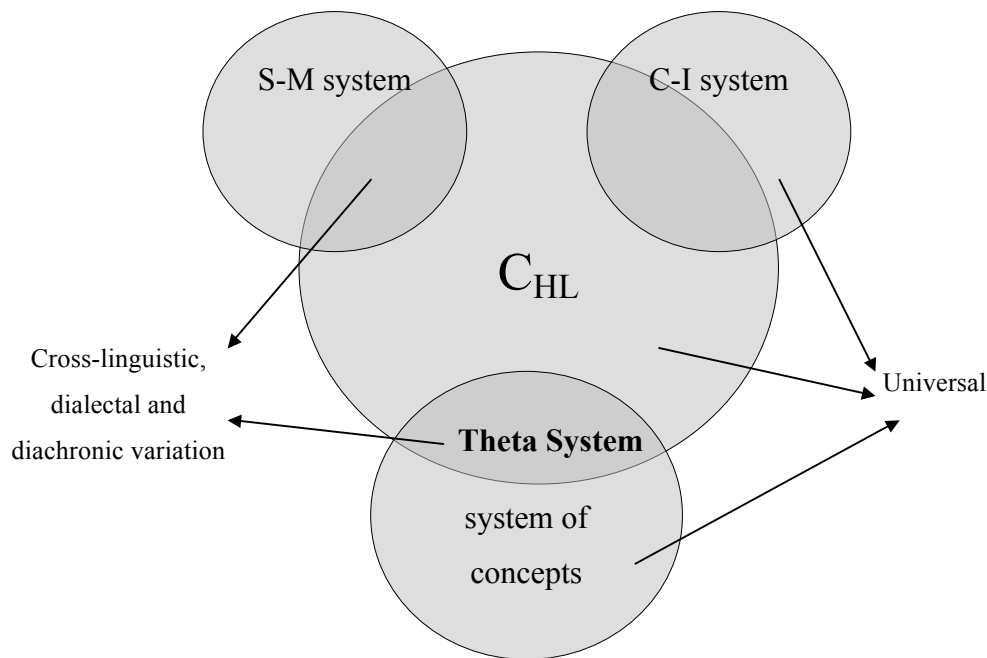
The theoretical framework upon which I will build the analysis is Principles and Parameters, and more concretely, I will follow the lines proposed by Chomsky's (1995, 2001, 2005) Minimalist Program. The language will be understood as a cognitive system that pairs sounds and meanings; in much the same way as Saussure (1916) defined *langue* as a mental system that associated sound images and concepts.

In (18) a graphic representation of the model of the Faculty of the Human Language (F_{HL}) can be seen. F_{HL} consists of a Computational System (C_{HL}) or syntactic component that combines lexical items taken from the System of Concepts by means of the Theta System (to be explained below). The C_{HL} combines the lexical items and the syntactic derivation proceeds by phases (Chomsky 2001, 2005). At one point, the derivation is transferred to the external systems through the linguistic interfaces: the Sensory-Motor System (S-M) that materializes the derivation (sound) and the Conceptual-Intentional System (C-I) that interprets the derivation (meaning).

As said before, if we are to maintain the C_{HL} and the C-I system universal (Chomsky 2005), as well as the System of Concepts (Reinhart 2002), we must place the linguistic variation (cross-linguistic, dialectal and diachronic) at the interfaces of the C_{HL} with the external systems.

(18) The Faculty of Human Language:

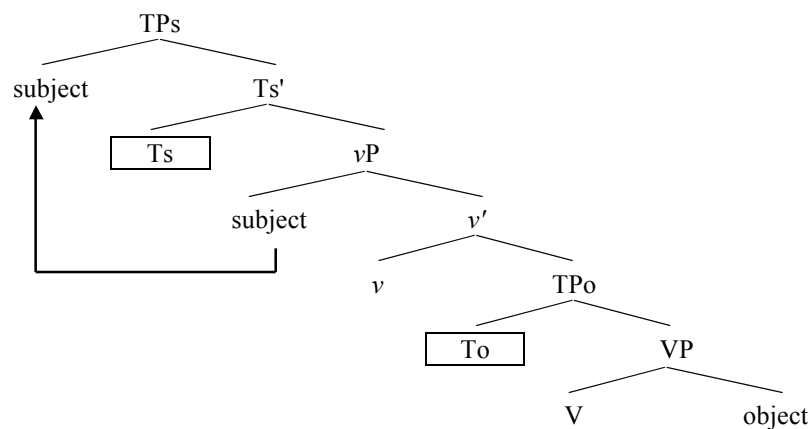
(based on Reinhart 2002)



I will pursue Chomsky's idea of a narrow syntax consisting of basic operations (*Agree* and *Internal/External-Merge*), and the interface between the syntax and the external systems, i.e. the Sensory-Motor (S-M) system, the Conceptual-Intentional (C-I) system and the lexicon.

Nonetheless, I will depart from Chomsky's view in several ways. As I will explain in chapter 5, I will follow Reinhart's (2002) conceptualization of the lexicon-syntax interface. This interface will be conceptualized as the Theta-System, which is the system that interfaces between the system of concepts and the syntax. It consists of lexical entries (the traditional "lexicon"), a set of marking procedures and a set of arity operations (i.e. the lexicon is an *active* module).

As for the structure of the verbal predication, I will follow Pesetsky & Torrego (2004) rather than Chomsky (2001). According to these authors, verbal predication, represented in (19), involves two subevents, each of which requires a tense node in syntax (there would be two tense nodes corresponding to the two subevents, even in states: Ts and To).

(19) Verbal predication structure:*(Pesetsky & Torrego 2004:503)*subject **T_s** [_{VP} *v* **T_o** [_{VP} V object]]

One tense node (T_s) licenses the external argument of the verb, i.e. the subject, whereas the other one (T_o) licenses the object. Pesetsky & Torrego argue that tense is introduced in the derivation from the lexicon in V. Since they do not detail which tense is exactly introduced by which element, I will assume that T_s is introduced by v whereas T_o is introduced by V. T features are interpreted in T_s and T_o respectively, i.e. both T_s and T_o have an interpretable but unvalued T feature whereas both v and V have an uninterpretable valued T feature. By means of a probe-goal relation, the interpretable T features of T_s and T_o get valued. On the other hand, DPs also have an uninterpretable T feature (this is what traditionally has been called structural Case), which is unvalued. This feature gets valued by agreement with either T_s , in the case of the subject, or T_o , in the case of the object.¹⁷

In table (20) the notation that I will be using throughout this work for functional heads, features and morphological realization of Case is summarized:

(20) Functional heads, features and Case - Notation:*(based on Pesetsky & Torrego 2004)*

Functional head	Feature	Morphological realization
T_s	Tns-s	NOM
T_o	Tns-o	ACC

I will also depart from Chomsky's conceptualization of feature valuation and deletion, summarized in (21), (22) and (23) below. I will follow, instead, that proposed by

¹⁷ See chapter 3 section 3.3.3. onwards for further discussion.

Pesetsky & Torrego (2007). They state that the operation *Agree* results in feature sharing, as formalized in (24). There are no uninterpretable features in a derivation but uninterpretable instances of a given feature. This means that a feature (occurrence) may have multiple instances. At least one of these instances must be interpretable so that the occurrence (the feature itself) can be interpreted by the semantic system. Uninterpretable instances are eventually "deleted".

(21) Agree (Assignment version): (following Chomsky 2001¹⁸)

- (i) An unvalued feature *F* (a *probe*) on a head *H* scans its c-command domain for another instance of *F* (a *goal*) with which to agree.
- (ii) If the goal has a value, its value is assigned as the value of the probe.

(22) Valuation/Interpretability Biconditional: (Chomsky 2001:5¹⁹)

A feature *F* is uninterpretable iff *F* is unvalued.

(23) Deletion of uninterpretable features: (Pesetsky & Torrego 2007:266)

Once an uninterpretable feature is valued, it can and must delete.

(24) Agree (feature sharing version): (Pesetsky & Torrego 2007:268)

- (i) An unvalued feature *F* (a *probe*) on a head *H* at syntactic location α ($F\alpha$) scans its c-command domain for another instance of *F* (a *goal*) at location β ($F\beta$) with which to agree.
- (ii) Replace $F\alpha$ with $F\beta$, so that the same feature is present in both locations.

Pesetsky & Torrego claim, again contra Chomsky (2001) that valuation and interpretation are independent. In other words, the elimination of the Valuation/Interpretability Biconditional in (22) allows lexical items to come from the lexicon with features that display to combinations of properties: first, uninterpretable but valued, and second, interpretable but unvalued. These cases are impossible for Chomsky's system due to the aforementioned Valuation/Interpretability Biconditional (22). Table (25) below shows the typology of lexical items that Pesetsky & Torrego's (2004) system allows.

¹⁸ Cited in Pesetsky & Torrego (2007:266).

¹⁹ Cited in Pesetsky & Torrego (2007:266).

(25) Types of features²⁰: (Pesetsky & Torrego 2007:269)

<i>u</i>F	val	uninterpretable, valued	<i>i</i> F	val	interpretable, valued
<i>u</i> F	[]	uninterpretable, unvalued	<i>i</i>F	[]	interpretable, unvalued

The notation in (26) will be used to refer to agree-chains throughout the thesis. Finally, I will distinguish three types of agree chains (27), depending on which kind of feature is shared:

(26) Agree-chains notation:

- a. $\alpha [uR[-]] \& \beta [iR[val]] \rightarrow Agree \rightarrow \alpha [uR[val]] \& \beta [iR[val]]$
- b. $R \{ \alpha, \beta \}$

(27) Types of agree-chains:

- a. **φ -chain**: agree-chain formed when two or more lexical items share one or more φ -features.
- b. **Tns-chain**: agree-chain formed when two or more lexical items share a Tns (tense) feature.
- c. **θ -chain**: agree chain formed when two or more lexical items share a θ -feature.

The operation Agree forms chains (**agree-chains**) between instances of a feature, whether or not those instances are valued (although at least one of them will eventually have to be valued so that the feature can be interpreted at the C-I system). These agree-chains are different from the relations established between a probe and a goal: two elements α and γ in such a configuration as (28) may be sharing a feature F and thus be part of an agree relation not due to a probe-goal relation between α and γ but because γ and β have a probe-goal relation and so have β and α . In this case, γ and α share the feature F, i.e. they are both instances of F but there is no probe-goal relation whatsoever between γ and α . Consequently, there cannot be movement of γ triggered by F on α because there is no probe-goal relation between α and γ , in spite of the fact that both share (are instances of) the feature F due to the two probe-goal relations established between γ and β , and between β and α . I will argue that this happens in cases of *indirect Exceptional Case Marking*, which will be introduced in chapter 3. In this sense, β intervenes between α and γ , preventing a probe-goal relation between them (though it

²⁰ Boldface = disallowed in Chomsky (2001).

Chapter 3

Control and null SE-anaphors at the interfaces¹

3.1. Introduction

In this chapter I will address the issue of control phenomena, as well as the nature, syntactic properties and the referential interpretation of the pronominal anaphor PRO.

Control phenomena, and their differences with raising constructions, have been one of the research topics in Generative Grammar since Rosenbaum (1967). On the one hand, raising constructions as (1), which are formed by verbs that do not assign an external theta role nor accusative Case, were considered the result of moving the subject from the embedded clause to the subject position in the matrix clause as in (1b).

- (1) a. D-STRUCTURE: \emptyset seem [John to be ill]
 b. S-STRUCTURE: John_i seems [t_i to be ill] (movement)
 c. "*John seems to be ill.*"

On the other hand, control constructions as (2) have been thought to be caused by a process that Rosenbaum called *Equi-NP Deletion*, which implies that one of two identical arguments at D-Structure is deleted at S-Structure as in (2b), so movement is no required.

- (2) a. D-STRUCTURE: John promised [John to pass the exam]
 b. S-STRUCTURE: John promised [~~John~~ to pass the exam] (*Equi-NP Deletion*)
 c. "*John promised to pass the exam.*"

Typical examples of control phenomena are given below: (3a) shows what is known as *obligatory control* (OC), in (3b) *non-obligatory control* (NOC) can be found. Finally, (3c) shows a null subject with arbitrary interpretation. Note that although (3c) seems not to be an instance of control, some authors (see Manzini & Roussou 2000:427) have pointed out that this kind of constructions might be the result of control by an arbitrary operator, as we will explain later on. Hence, I will consider (3c) as the third main type of control: *arbitrary control* (AC).

¹ I wrote an earlier version of this chapter when I was studying at UiL-OTS (University of Utrecht, The Netherlands) under the supervision of Eric J. Reuland.

- (3) a. John wants to read The Silmarillion.
b. Reading the Silmarillion is difficult for me.
c. Eating before going to bed is not good.

In the framework of Government & Binding (G&B, Chomsky 1981) the analysis of control phenomena was based on the null element PRO, which was analyzed as a pronominal anaphor. The two types of control (OC and NOC), as well as arbitrary constructions were considered a result of the presence of PRO as represented in (3'):

- (3') a. John wants [**PRO** to read The Silmarillion].
b. [**PRO** Reading the Silmarillion] is difficult for me.
c. [**PRO**_{arb} Eating before going to bed] is not good.

Its distribution was derived from the way in which its properties interact with the Binding Theory. As a pronominal, PRO has to be free in its governing category, and as an anaphor it has to be bound. This apparent paradox is resolved if PRO has no governing category. Hence PRO is only able to appear in ungoverned positions. This statement, which is known as the PRO Theorem, is in (4) below.

- (4) PRO Theorem: *(Chomsky 1981:191)*
PRO must be ungoverned.

In ungoverned positions, PRO cannot bear Case², and hence its lack of phonological content³. Being exempt from the binding theory, its interpretation was handled by Control Theory, which was a distinct module of the Grammar.

In Chomsky (1982), a contextual definition of empty categories was introduced. PRO was an empty nominal element indistinct in the lexicon from the other null categories, A'-trace, A-trace and *pro*. It was proposed that the manner in which they were governed and bound determined their eventual referentiality.

Although this theory of empty categories was elegant and covered much of the empirical data at that time, more extensive research on control configurations across languages showed that the issue is far more complex.

One theoretical shortcoming is that Control Theory is based upon Rosenbaum's (1967) Minimal Distance Principle (MDP), which states that PRO is controlled by the

² In this state of the theory, Case was assigned under government.

³ In G&B, the Case Filter is a phonological filter: if an argument is to be visible at P.F., it has to be Case-marked. PRO, being not Case-marked, is invisible at P.F. and hence, it is not pronounced.

closest available antecedent. However, the MDP could not be easily reduced to some independently needed principle in the Grammar in newer stages of the theory, something that is crucial in the Minimalist Program (Chomsky 1995). In section 3.3.2.1 I will review Manzini's (1983) proposal of Control Theory as derived from PRO being a pure anaphor. She claimed that its distribution should be determined by its lack of Case, rather than the PRO Theorem. Hence, she reformulated Chomsky's (1981) Binding Theory, in particular Condition A, in order to cover the distribution of pronominals, anaphors and PRO (which she considered to be a pure anaphor). However, it is far from settled that PRO is Case-less. Plenty of evidence has been presented that supports the claim that PRO bears Case, both inherent and structural (see Andrews 1971,1976; Comorovsky 1985, Sigurdsson 1991, Landau 2006). In conclusion, Control Theory and PRO have been rather controversial in the field of syntax and semantics. Neither the interpretation of PRO nor its distribution is easily integrated in the theories so far proposed.

In this chapter I present a novel analysis of control and PRO based on Teomi (2005) and integrated in the Minimalist Program (Chomsky 1995 onwards). It is organized as follows: in section 3.2, I will address the problems raised by current approaches to control, namely the Null Case approach, Hornstein's (1999) Movement Theory of Control, and Landau's (2004b) Calculus of Control. In section 3.3 I will introduce the theoretical background upon which the analysis is built. In section 3.4 I will develop the analysis, which sets out to derive the interpretation of PRO from Binding Theory as conceived by Reuland (2001, 2006, 2008), and its distribution from the interaction of Case and Phase theories. One of the ultimate goals is to unify the analyses of PRO and *pro* by conceiving both empty categories as (null) SE-anaphors in section 3.7.1. Evidence from English *ing*-clauses and Spanish infinitives will be presented in sections 3.5 and 3.6 respectively in order to support the approach to PRO defended in this thesis. I will present in section 3.7 some notes on several issues that might be interesting for future lines of research. Finally, the conclusions will be presented in section 3.8.

3.2. Current approaches to control

The most serious theoretical shortcoming of the binding-based approach to the distribution of PRO within the G&B framework came along with the Minimalist Program (MP) introduced by Chomsky (1995). The MP sets out to explain the

properties of the human language on the basis of more fundamental notions. Chomsky (1995) claims that concepts such as *government* are amenable to explanation in more primitive terms themselves. Hence any result based on the notion of *government* should be reassessed. Since there is no direct reconstruction of the notion of *government* and *governing category* within the MP, PRO had to be reanalyzed without resorting to these notions, and hence, the PRO Theorem in (4) is no longer useful.

There is an intense debate about control and PRO nowadays. This debate focuses mainly on Hornstein's (1999) Movement Theory of Control and Landau's (2004b) Calculus of control (c.f. Boeckx & Hornstein 2004, Hornstein 2003, Landau 2003, Landau 2004a). All the approaches to be reviewed in this section suffer from empirical and theoretical shortcomings. They rely on either theoretical stipulations (like the R-rule or the notion of null Case) or controversial theoretical positions (movement into theta positions).

This section discusses two minimalist approaches to control and PRO, Hornstein's (1999) Movement Theory of Control (MTC) and Landau's (2004b) Calculus of Control, as well as Chomsky & Lasnik's (1993) Null Case Theory,

3.2.1. The Null Case Theory

Since the notion of government was no longer available in MP, Chomsky & Lasnik (1993) proposed to go back to the idea of explaining the distribution of PRO by means of Case Theory. This led them to claim that PRO, like any other DP, needs to be assigned Case (its Case feature must be checked) but it can only bear a special kind of Case: Null Case, which, in turn, can only be assigned/checked by infinitival T° , i.e. the tense head present in English *to*- and *ing*- infinitivals. Although this theoretical step allows dispensing with the notion of government, it is rather stipulative and generated many theoretical and empirical problems, as Hornstein (1999) and Manzini & Roussou (2000) state:

- a) The Null Case approach basically stipulates the distribution of PRO. Null Case is special in two ways: first, it is designed to fit only one expression: PRO. And second, only non-finite T can assign/check it. Since null Case and PRO are always seen together, there is no independent way of establishing the existence of either. Hence, null Case seems to be an ad-hoc description of the positions that PRO occupies rather than a genuine explanation for the distribution of PRO. Manzini & Roussou (2000) state that this is a step back with respect to the G&B framework,

where the distribution of PRO did indeed follow from the interaction of assumptions that were independently motivated.

- b) Even if null Case can account for the distribution of PRO, Control Theory is still to be explained. In other words, the referential interpretation of PRO (whether it is controlled by a local or non-local antecedent, or it is non-controlled at all) still has to be accounted for.
- c) Case marked empty categories, such as A'-trace, block contraction like in the case of *wanna* in (5b). Null Case marked PRO fails to block such a contraction in (5a).
- (5) a. Who do you want [*wh*-trace to vanish]? (Hornstein 1999:75)
 *Who do you wanna vanish?
- b. John's going [NP-trace to leave]. (Hornstein 1999:75)
 John's gonna leave.

Furthermore, environments where both null and overt subjects are allowed pose an empirical problem for approaches like the Null Case Theory (but also Hornstein's Movement Theory of Control in section 3.2.2.1, and Manzini & Roussou's approach in section 3.2.2.4), which tries to account for the distribution of PRO as complementary to overt subjects. Two such environments are English *ing*-clauses and certain Spanish infinitivals:

- a) English *ing*-clauses can surface either with PRO or a lexical DP (mostly in accusative) in subject position, as shown in (6a) and (6b) respectively. If infinitival *ing* is a null Case assigner, an accusative subject is unexpected. Infinitival *ing* cannot check the accusative or nominative Case of the embedded subject, only the null Case of PRO. Therefore a nominal other than PRO should not be licensed unless there is some kind of ECM, which seems not to be the case (at least ECM like in *believe*-type of verbs) as the subject of the *ing*-clause cannot be moved into the matrix clause by means of A-movement (Reuland 1983), as shown in (6e), unlike in the case of *believe*-type verbs in (6d).
- (6) a. I love [**PRO** singing that song]
 b. I love [**Mary** singing that song]
 c. I expect [**Mary** to sing that song]
 d. **Mary** is expected [~~Mary~~ to sing that song]
 e. ***Mary** is loved [~~Mary~~ singing that song]

b) Spanish does not usually allow lexical subjects in infinitive clauses when these are selected by transitive verbs, unlike English *ing*-clauses that have this option.

(7) a. Yo quiero [**PRO** leer el libro]

I want [**PRO** read_{infinitive} the book]

"I want to read the book"

b. *Yo quiero [**María** leer el libro]

I want [**María** read_{infinitive} the book]

"I want Maria to read the book"

Nonetheless, Spanish infinitivals may surface with nominative subjects (Nominativus-cum-Infinitivo, NcI) in certain syntactic environments (Fernández Lagunilla 1987; Rigau 1993,1995; among others):

a) When the infinitive clause is a complement to a preposition:

(8) Abandoné la casa antes de volver Pedro

pro Left_{1st,sing} the house before of come back Pedro

"I left the house before Pedro came back"

b) When the infinitive occupies the subject position in the matrix clause:

(9) Ir yo a la universidad mañana será difícil

Go_{inf} I to the university tomorrow will be difficult

"It will be difficult for me to go to the university tomorrow"

c) In exclamative and interrogative infinitival clauses:

(10) ¿Ser yo tonto?

Be_{inf} I stupid?

"Me stupid?"

If Spanish infinitives can only check null Case, as English infinitives are supposed to do, this can account for the contrast in (7). However, sentences in (8), (9), and (10), tell us that this is not always the case and the infinitives seem to be able to check the nominative Case of the overt subject as well (or any other Case different to Null Case, which can only be assigned to PRO). I will come back to this issue in section 3.6.1.

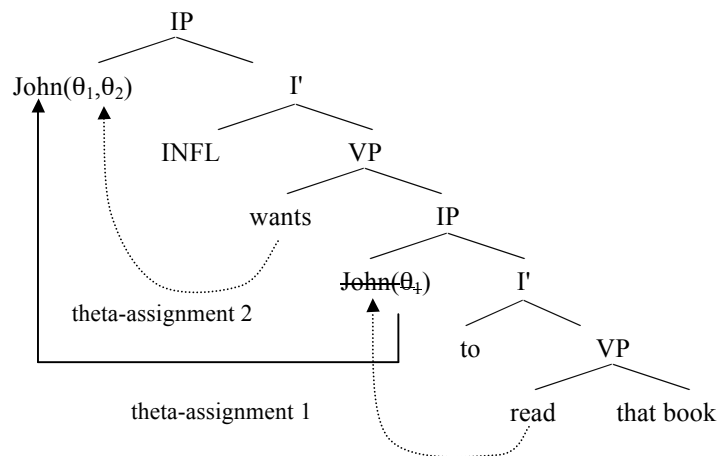
3.2.2. The Movement Theory of Control

3.2.2.1. Hornstein (1999)

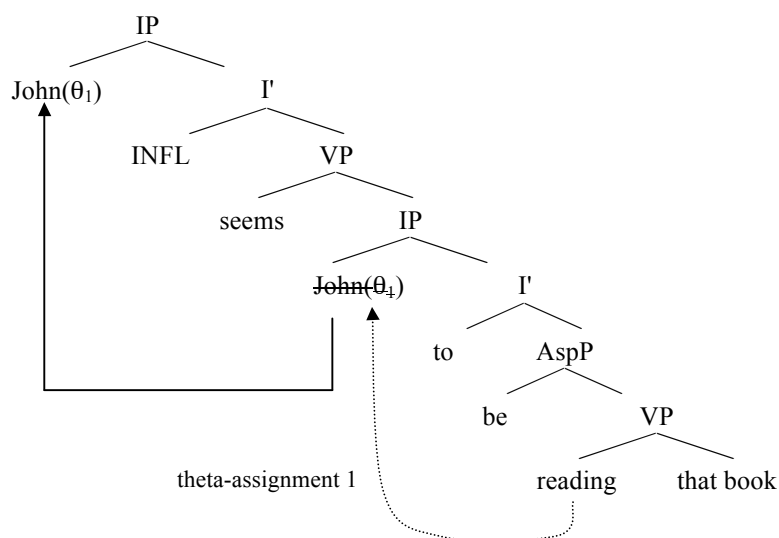
Hornstein aims to dispose of both Control Theory and the notion of PRO by reducing OC to A-movement effects, and PRO to a trace left by an NP that has undergone A-movement. In order to do this, the Theta-criterion must be dispensed with. Hornstein states that this step is not only possible but also a necessary one in the MP as it is a redundant notion once D-Structure is eliminated as an independent level of representation from syntactic theory, as the MP does. The consequences are in principle twofold: the ban on one theta role per DP is abandoned, and consequently, movement into theta-positions is permitted if theta roles are analyzed as features that need to be checked.

Control and raising structures are both analyzed as the result of A-movement. The only difference between the two is that in raising, the embedded subject moves to a non-theta-position (the embedded subject checks only one theta feature), whereas in control structures the embedded subject moves into a theta-position and it checks two theta features.

(11) a. John_{θ₁,θ₂} wants John_{θ₁} to read that book.



b. John_{θ₁} seems John_{θ₂} to be reading that book.



The distribution of PRO now follows from Case: PRO is the unpronounced copy of a nominal that moves to get its Case feature checked. Hence, PRO itself lacks Case.

Control Theory can be abandoned as a separate module of the grammar since the semantic properties of OC are derived from the effects of A-movement and the MDP can be now reduced to a condition postulated for independent reasons, namely the Minimal Link Condition (MLC) that applies to movement (Chomsky 1995). Subject control in verbs like *promise* follows from the MLC, just as object control does, by saying that the object is the complement of a null preposition.

NOC cannot be unified with A-movement and is considered in MTC as an instance of *pro*. It shows up in syntactic islands, i.e. when movement is blocked. Hornstein argues that it is a last-resort mechanism to save the derivation, and compares NOC with the phenomenon of *do*-support in English.

Finally, OC in adjuncts can be accounted for by means of *sideward movement*⁴ (Boeckx & Hornstein 2004), defined in (12): in a configuration like (13a) it is not clear why PRO can be controlled by NP₁ and not by NP₂. One could state that NP₂ never c-commands the adjunct and so, PRO can never be bound by NP₂. However, Hornstein argues that (13b) shows that *every book* c-commands *it* at LF because *every book* can bind *it*. So the point is why NP₂ cannot control PRO if it c-commands PRO at LF.

⁴ Hornstein emphasizes that although this mechanism helps in accounting for the cases of OC in adjuncts, it is not indispensable for the MTC.

(12) Sideward movement: (Boeckx & Hornstein 2004:437)

Movement taking place between two subtrees that are not connected at the time of movement.

(13) a. NP₁ V NP₂ [_{adjunct} PRO_{1/*2} ...]

b. John_i reviewed every book_j [without PRO_{i/*j} reading it_{*i/j}] (Hornstein 1999:88)

The solution provided by Hornstein is that at some point of the derivation, NP₂ and PRO are equidistant from matrix T (since at that point, none c-commands the other). PRO (i.e. NP₁) sideward-moves (i.e. from one tree to the other) to the matrix [Spec, vP] (before the vP and the adjunct merge) receiving a second theta-role. It further raises to [Spec,TP] in order to get its Case feature checked. The option of NP₂ moving is disallowed because in such a situation, NP₁ would not check its Case and so the derivation would crash.

3.2.2.2. *Manzini & Roussou (2000)*

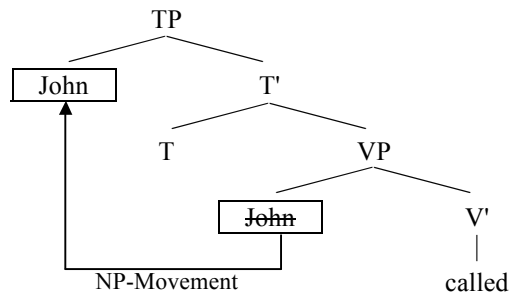
Manzini & Roussou also aim to reduce control to A-movement, though their ultimate goal is to describe A-movement in terms of Move-F (Chomsky 1995) instead of phrasal-movement, composed of copy and merge, which has been proved appropriate to characterize A'-movement.

They assume, along with Hornstein (1999), that the Theta Criterion must be disposed of along with the notion of D-Structure. However, they note that the consequences of this theoretical step are more crucial than just reducing control to A-movement but may help to obtain a more accurate characterization of A-movement itself. They argue that there is no evidence either at LF (reconstruction effects) or at PF (effects on phonosyntactic rules on PF like *wanna*-contraction) that supports the claim that phrasal movement is involved in A-movement.

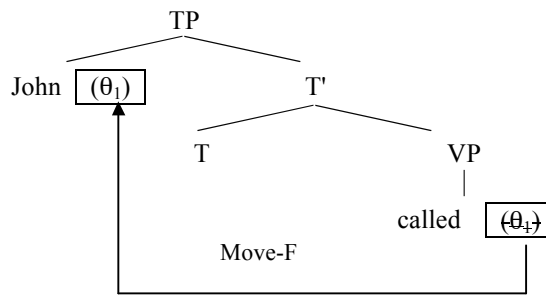
Their claim is that lexical items are generated in the position where they receive Case. This eliminates the redundancy of using both notions of structural Case and D-feature in the system of Chomsky (1995). Instead of phrasal movement, movement of features (Move-F) connects the lexical item with a theta feature associated with the predicate. This is exemplified in (14): *John* merges in the specifier of the vP in (14a) and gets its theta role. After that, it moves to [Spec, TP] in order to check its Case feature and the D-feature of T. On the other hand, in (14b) *John* merges directly in [Spec, TP] and there it attracts the theta-role of the predicate, which is understood as a

feature that must be checked by moving to *John*. Manzini & Roussou also claim that it is the interpretable D-feature of the DP the element that attracts the theta-feature in the predicate.

(14) a. Phrasal movement:

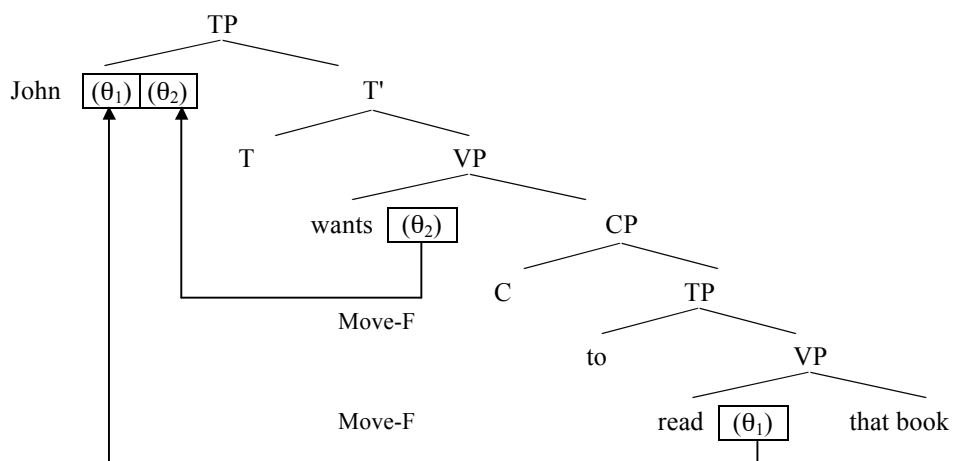


b. Feature movement (Move-F):

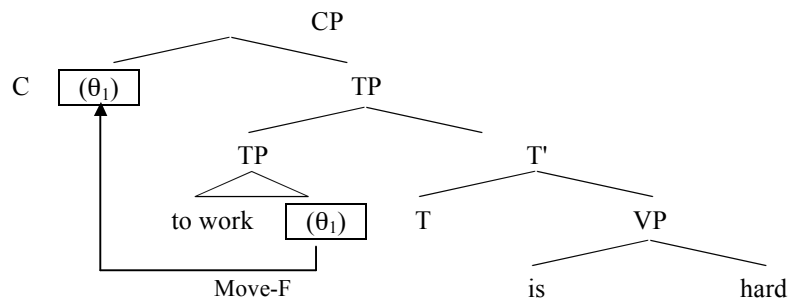


In this framework, control can be construed as a configuration in which a single DP attracts more than one predicate, i.e. two different theta-features move to one DP, as in (15).

(15) $John_{\theta_1, \theta_2}$ $wants_{\theta_1}$ [to $read_{\theta_2}$ that book]



Arbitrary control is reduced to the attraction of a predicate by an operator in C:

(16) [C_{θ1} [it is hard [to work_{θ1}]]]

Finally, basic locality properties of control follow from a modified scopal version of Chomsky's (1995) MLC and from Kayne's (1984) notion of Connectedness, both phrased as conditions on the attract operation, which precedes Move-F in the system of Chomsky (1995).

3.2.2.3. Theoretical problems

The MTC is part of a more general approach to dependencies that Safir (2004) labels *Movement As Co-construal (MAC)* theory. The advantage of reducing co-construals to movement is, on the one hand, that the c-command condition on bound readings, typical of anaphors, can be derived from the fact that movement requires c-command. On the other hand, with the introduction of *sideward movement*, the distribution of the movement relation can provide a more accurate guide to the possible relations between co-construed nominals than c-command, because it derives exactly those cases where c-command fails to license a bound reading, though the bound reading exists.

Safir notes, however, that in order for Principle C to be derived from the impossibility of backwards movement, all co-construals need to be reduced to movement. The consequence is that movement must now be presumed to relate positions not only where c-command does not hold, but also across all sorts of syntactic islands.

Besides, movement must also be allowed to take place between positions in different sentences, as in (17).

(17) *Frodo* is near Rivendell. *He* has the One Ring.

The problem is that there are pronouns that have no linguistic antecedent, which must be allowed to be generated for the MAC to work.

To conclude, reducing all co-construals (control among them) to movement requires thinking of movement as a new more complex operation that applies beyond c-

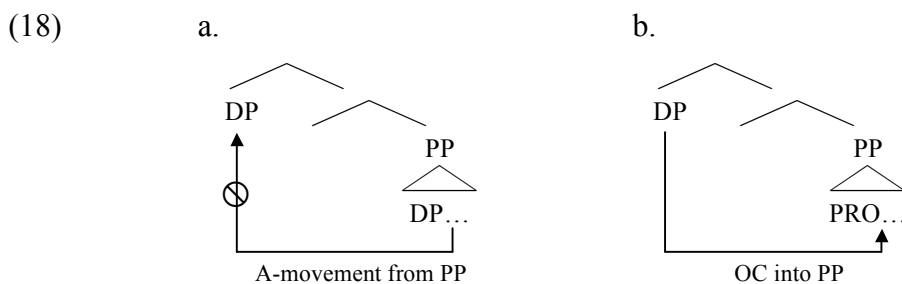
command, across syntactic islands, and even between different sentences. This theoretical step has notable disadvantages that are more likely to obscure the interesting issues rather than resolve them.

Finally, the elimination of the Theta Criterion is quite controversial and many authors do not accept this claim (Chomsky 2001, 2005; Landau 2004, 2006; Reuland 2001, among many others).

3.2.2.4. Empirical problems

English *ing*-clauses allow, as noted in the previous section, both overt and null subjects. This is in principle unexpected under the MTC since this would indicate that the subject position of these clauses is a position from whence movement is both possible and impossible. Pires (2001) offers a solution to this puzzle under the assumptions of the MTC (see section 3.5.1). In section 3.4 I will provide an alternative analysis that does not rely on the theoretical assumptions that the MTC requires.

The case of prepositions selecting infinitives with overt subjects in Spanish also poses a problem for the MTC as OC PRO is predicted to be allowed only in those positions from whence movement can take place. However, movement from PPs is not allowed in Spanish, as can be shown in (19) (preposition stranding yields ungrammatical results in Spanish).



- (19) a. Juan_i abandonó la casa [PP después de decir PRO_i la verdad]
 John_i left the house [PP after of say_{inf} PRO_i the truth]
"John left the house after saying the truth"
- b. *¿Qué_j abandonó Juan_i la casa [PP después de decir PRO_i t_j]?
 What_j left Juan_i the house [PP after of say_{inf} PRO_i t_j]?
 **"What did Juan left the house after saying?"*

If movement from PPs is banned in Spanish, it is unexpected that PRO is allowed there if it were the result of movement. However, agreement from within a PP is possible.

This can be seen in PPs where the SE-anaphor *si*⁵ is bound by a DP from outside the PP, as in sentence (20) below. If PRO is not the result of A-movement but a SE-anaphor, this contrast can be easily accounted for. In fact, this is the line of argumentation that I will take in section 3.4.

- (20) Juan_i se compró el libro [PP para **si**_i?(mismo)]
 Juan_i *se* bought the book [PP for **si**_i?(zelf)]
 “*Juan bought the book for himself*”

Finally, Culicover & Jackendoff (2001) point out more empirical problems regarding the Movement Theory of Control. Among other things, they show that semantic characteristics of the predicates are relevant for the establishment of the controller of OC PRO in certain cases. This cannot be accounted for by an exclusively syntactic theory of control as Hornstein's.

3.2.3. The Calculus of Control

Landau (2004b) argues that OC environments do not form a natural category but OC is the "elsewhere" case of the environments where overt subjects can emerge. The claim is that PRO is a null SE-anaphor (in the sense of Reinhart & Reuland 1993). Its interpretation follows from the operation Agree (Chomsky 2001), and its distribution from the interaction of Agr (φ -features) and T(ense) features both on I and C. The formal rules presented in (21) and (22) control the assignment of values to the Agr & T features on I and C.

- (21) Specifying [T] on embedded I/C: (Landau 2004b:839)

- a. Anaphoric tense → [-T] on I/C
- b. Dependent tense → [+T] on I/C
- c. Independent tense → [+T] on I, Ø on C

- (22) Specifying [Agr] on embedded I/C: (Landau 2004b:840)

- a. On I:
 - i) overt agreement → [+Agr]
 - ii) abstract agreement → [-Agr]
 - iii) no agreement → Ø

⁵ In chapter 4 I will define *si* as a SE-anaphor like Dutch *zich*. I will state that it occupies A-positions, and it also checks structural case, hence it can be the complement of prepositions.

- b. On C: i) [+Agr] → [+T]
 ii) otherwise → ∅

Landau's R-rule accounts for the interaction of these features with the licensing of PRO (see ((23))). It basically states that whenever I and C are specified for [+T,+Agr], then they automatically come to bear [+R], which is an interpretable feature on nominals and indicates that they can refer to objects in the world without an antecedent, whereas [-R] implies that the nominal requires an antecedent in order to be able to be referential (Reinhart & Reuland 1993). Note that [±R] is interpretable on DPs and uninterpretable on both I and C. Any other feature constitution, i.e. [+T,-Agr], [-T,+Agr+],[-T.-Agr], is associated with [-R]. Notice that lack of [T] or [Agr] renders the rule inapplicable, i.e. no [R] value is assigned. DP/*pro* are [+R], hence they can delete the [+R] feature on I/C, whereas PRO is [-R]⁶ and it is required when I/C have an uninterpretable [-R] feature.

(23) R-assignment Rule: (Landau 2004b:842)

- For $X_{[\alpha T, \beta Agr]} \in \{I, C, \dots\}$:
 $\emptyset \rightarrow [+R]/X_{[_]}$, if $\alpha = \beta = '+'$
 $\emptyset \rightarrow [-R]/\text{elsewhere}$

In this system, OC is predicted not only in infinitives but also in other environments such as Hebrew finite embedded clauses and Balkan subjunctives. I will illustrate the case of infinitives and Hebrew finite control.

English infinitives are thought to have an I marked as [-T-Agr] since its tense is anaphoric, this is to say that the clause has no tense specification but "shares" the one of the matrix clause, and there is no morphological agreement. The C head is marked as [-T], due to the anaphoric nature of the clausal tense and hence it has no [R] feature. Since I has an uninterpretable [-R] feature, PRO with interpretable [-R] is required in order to delete it.

(24) I [-T, -Agr] → [-R] (English infinitive clauses; Landau 2004:847)
 C [-T]

Hebrew is a partial Null Subject Language (NSL). It allows null 1st and 2nd person pronouns but not 3rd person pronoun dropping. There is however one environment

⁶ Whereas Reinhart & Reuland (1993) suggest that PRO is [+R], Landau assumes that it is (at least in OC environments) [-R], which means that it needs an antecedent in order to be interpreted.

where a null 3rd person pronoun appears: in subject position of embedded subjunctive clauses, as in (25).

- (25) Hem_i kivu se-hem_{i/j}/pro_{i/*j} yelxu ha-bayta mukdam. (Landau 2004b:816)
 They hoped that-they/*pro* will-go_{3rd} home early
 They hope that they would go home early
 "They hoped to go home early"

- (26) I [+T, +Agr] → [+R] (Hebrew subjunctive clauses; Landau 2004b:865-866)
 C [+T, +Agr] → [+R]

Landau analyzes this pronoun as an instance of PRO instead of *pro* and argues that the Calculus can account for this: the embedded subjunctive clause gets its I marked [+T+Agr] as it has dependent (irrealis) tense and overt agreement. C gets also marked [+T+Agr] due to its dependent Tense. In principle, these clauses need to delete the uninterpretable [+R] features on I/C by means of an overt DP with interpretable [+R]. However, both uninterpretable [+R] features on I and C can check off each other. This makes the presence of PRO possible since there is no uninterpretable [R] feature to delete either on C or I.

It is worth emphasizing that Landau (2005) follows authors like Andrews (1971, 1976), Comorovsky (1985), and Sigurdsson (1991) and argues that PRO bears standard Case rather than a special kind of Case like the null Case proposed by Chomsky & Lasnik (1993).

Landau's Calculus of Control has the advantage of making it very explicit how the distribution of PRO follows and it reduces Control Theory to the operation Agree available in UG. However, it does not explain why the distribution of PRO should be dependent on the presence of T and Agr. As it is stated now, the Calculus is a description rather than an explanation. The R-rule, on the other hand, is stipulative (as Landau himself recognizes) as well as the R feature⁷. Finally, some empirical problems indicate that a refinement of the Calculus is needed⁸:

⁷ Reuland (2001) develops a theory of binding without resorting to referentiality features (the φ -features that the pronominals and anaphors possess, and the way they interact with the grammatical system is enough to account for their interpretation) such as the notions *anaphor* or *pronominal* used as primitives in the framework of Government & Binding, or the [R] feature present in the system of Reinhart & Reuland (1993). See section 3.3.3 for more details.

⁸ Actually, these empirical problems are shared by all the theories so far mentioned.

- a) **English *ing*-clauses** are problematic since their feature composition requires a PRO subject: if these gerunds are analyzed as having dependent tense (at least in some cases) and no overt agreement (Landau 2004a), this yields the feature configuration in (27), which requires a PRO subject in order to delete the uninterpretable [-R] feature on I. However, this need not be the case as gerunds in English can surface with overt subjects, as we saw in sentences like (6b), which is unexpected for the Calculus.
- b) Something similar happens with **Spanish infinitivals**, which allow overt subjects in certain syntactic configurations, as shown in (8), (9), and (10). Spanish infinitival complements to prepositions are temporally dependent and have no overt agreement; thus their feature composition, represented in (28), makes them require a PRO subject in order to delete the uninterpretable [-R] feature on I. The feature composition of these kinds of clauses is similar to English *ing*-clauses and the infinitive clauses that allow Partial Control. However, the latter do not allow overt subjects due to [-R] on I°, unlike both English *ing*-clauses and Spanish P+infinitive constructions.
- c) **Latin has temporal specified infinitives** (Cecchetto & Oniga 2001, 2004) that, on the one hand, allow overt subjects, and, on the other hand, disallow PRO in subject position. This is difficult to account for since the feature composition of these clauses is similar to Spanish P+infinitive and English *ing*-clauses (their tense specification is dependent on the matrix clause and there is no overt agreement), i.e. I° has a [-R] feature and C° has a [+R] feature. How is the uninterpretable [-R] feature deleted if an overt subject emerges? And why should be PRO prohibited in such a configuration?

- (27) I [+T, -Agr] → [-R] *(English ing-clauses)*
 C [+T, +Agr] → [+R]
- (28) I [+T, -Agr] → [-R] *(Spanish P+infinitive construction)*
 C [+T, +Agr] → [+R]
- (29) I [+T, -Agr] → [-R] *(Latin temporally specified infinitives)*
 C [+T, +Agr] → [+R]

3.2.4. Recapitulation

As said before, debate about control nowadays focuses on Hornstein's MTC and Landau's Calculus of control (Boeckx & Hornstein 2004, Hornstein 2003, Landau 2003, Landau 2004a). All the approaches reviewed in this section suffer from empirical and theoretical shortcomings. They rely on either theoretical stipulations (like the R-rule or the notion of null Case) or controversial theoretical positions (movement into theta positions).

The analysis that I put forward in this chapter aims to solve these problems. I will assume along with Landau, that PRO is a Case-marked null SE-anaphor in the sense of Reuland (2001) and Reinhart & Reuland (1993). I will argue that under that assumption, the Calculus and the R-rule are no longer needed in order to account for the distribution of PRO, which follows now in a natural way from the interaction of the Case necessities of PRO, and Phase Theory. The interpretation of PRO, on the other hand, is derived from Binding Theory as developed by Reuland (2001, 2006a, 2008).

3.3. Theoretical background

In this section, I will introduce the notions of anaphora, anaphora resolution and binding. I will also explain the Binding Theory both in its original version and in the more recent version I will use throughout this thesis. These are the theoretical tools upon which the analysis to be developed in section 3.4 will be build.

First, I will review the notions of deixis, reference and anaphora resolution, as well as Chomsky's (1981) Binding Theory. Subsequently, I will discuss the main shortcomings that Chomsky's Binding Theory has had to face. I will move on to discuss Reuland's (2001, 2006a, 2008) version of the Binding Theory. Finally, I will introduce the notion of agree-chain, which will be crucial for the analysis of control that I will introduce later on.

3.3.1. Anaphors and anaphora resolution

Nominal items refer to entities in a possible world. In this section I explore how the nominal elements (both "full" nominals and pronouns) get their reference by means of two phenomena: deixis and anaphora. I will also introduce Chomsky's (1981) Binding Theory and the three kinds of nominal items that it defines: pronominals, anaphors and R-expressions.

3.3.1.1. *Introduction: Deixis and anaphora*

Deixis is the phenomenon of understanding the meaning of certain words and phrases in an utterance, which requires contextual information. Deictic words (*deictics*) have a denotation that changes depending on the place, time or persons involved in the speech situation. Fillmore (1971) distinguished three major grammaticalized types of deixis:

- a. **Place or space deixis** refers to the spatial locations relevant to an utterance. The locations can be those of the speaker and addressee, or those of persons or objects that are referred to in the discourse. Examples of place deictics are place adverbs like *here* and *there*, as well as demonstratives like *this* and *that*.
- b. **Time deixis** is concerned with the various times involved in, and referred to in an utterance. Time deictics are time adverbs like *now*, *then*, *soon*, and *today*, as well as the different tenses (present, past, future, etc.)⁹.
- c. **Person deixis** is concerned with the grammatical persons involved in an utterance, both those directly involved (speaker and addressee) and not directly involved¹⁰ (overhearers, mentioned, etc.). This is generally accomplished with nouns like *Peter*, *Lucie*, etc. as well as with personal pronouns like *I*, *you*, *he*, etc.

Anaphora consists of an expression that refers to another expression, which can be either present in other utterances, or be an entity in the real (or a possible) world. According to Reinhart (1999) "*Anaphora is used most commonly in theoretical linguistics to denote any case where two nominal expressions are assigned the same referential value or range*". Halliday & Hassan (1989) distinguish two main types of anaphoric ties within a different framework:

⁹ Note that not all the tenses are deictic (see the classification of tenses in Bello (1841)). Simple past tense in English is a deictic tense in the sense that it denotes a time previous with respect to the speech time (which roughly corresponds with the deictic *now*). However, the past perfect is an anaphoric time that is previous with respect to some other past time rather than the speech time, which has to be defined in a previous utterance (see the analysis of English tenses of Reichenbach's (1947)).

¹⁰ Note, however, that I will defend later on that personal pronouns that refer to people not directly involved, i.e. third person pronouns such as *he* and *she*, are actually exophoras rather than deictics.

- a. **Exophora**¹¹ is an expression whose referent does not appear in the utterances of the speaker, but instead in the real world (i.e. the extra-linguistic context). In (30), *I* is a deictic that refers to the speaker, whereas *The Silmarillion* is an exophoric expression that refers to an entity (a book called "The Silmarillion") in the real (or a possible) world.
- b. **Endophora** is an expression whose referent is present in the utterances of the speaker, and we can classify them in two types:
- **Anaphora** if the referent is present in preceding utterances. In (31), *I* is a deictic, *The Silmarillion* is an exophoric expression and the pronoun *it* is an endophoric anaphora because it refers to *The Silmarillion* in the previous utterance.
 - **Cataphora** if the referent is present in forward utterances. In (32), the pronoun *it* is an endophoric cataphora since its referent is *The Silmarillion*, which is present in the forward utterance.

(30) I read The Silmarillion when I was young.

(31) a. I read The Silmarillion when I was young.

b. I didn't like it.

(32) a. SPEAKER A: I didn't like it.

b. SPEAKER B: What?

c. SPEAKER A: The Silmarillion, which I read when I was young.

3.3.1.2. *Strategies of anaphora resolution*

Reinhart (2000, 2006) proposes two strategies of anaphora (endophora and exophora) resolution: binding and valuation. For the time being, I will restrict the discussion to non-reflexive pronouns such as *he* and *she* (pronominals), and leave reflexive pronouns like *himself* and *herself* (anaphors), for section 3.3.1.3. Let us look at (33) and (34) to illustrate how an anaphora like the pronoun *she* can be resolved, i.e. get a referent¹² (there are similar examples in Spanish, see Otero's (1999:1235) examples (4)-(6)):

¹¹ In Generative Grammar, the notion of *referentiality* stands for exophora, which has been formalized in several ways; see R-expressions and Condition C of the Binding Theory (Chomsky 1981) or the $[\pm R]$ feature (referential independence) of Reinhart & Reuland (1993).

¹² The examples of this section are from Reinhart (2000), unless otherwise indicated.

(33) Lucie didn't show up today. (Reinhart 2006:165)

(34) Lili thinks she's gotten the flu. (Reinhart 2006:165)

The pronoun *she* in (34) can refer either to *Lili* or to *Lucie*. This is due to the fact that the anaphora (the pronoun *she*) can be resolved by resorting to the two aforementioned strategies: either binding or valuation.

(35) a. Lili (λx (x thinks z has gotten the flu)) (Reinhart 2006:165)

b. BINDING: Lili (λx (x thinks x has got the flu))

c. VALUATION: Lili (λx (x thinks z has got the flu) & $z = \text{Lucie}$)

d. COVALUATION: Lili (λx (x thinks z has got the flu) & $z = \text{Lili}$)

If *Lili* binds the pronoun, *she* obligatory refers to its antecedent *Lili*. Binding in (35b) implies the closure of the VP property at the semantic level by binding the variable of the pronoun by one λ (lambda) operator. Reinhart proposes to understand binding in terms of logical syntax. In other words, binding (whether we refer to argument binding or to operator-variable binding) always implies a relation between operators and variables¹³. For the time being, I will take for granted that the logical-syntax definition of binding that Reinhart proposes in (36) is preferable in order to understand the two anaphora construals (35b) and (35c-d) that arise from the sentence (34).

(36) A-Binding (logical-syntax based definition): (Reinhart 2006:166)

α A-binds β iff α is the sister of a λ -predicate whose operator binds β .

Hence, *Lili* A-binds the pronoun *she* because it is the sister of the λ -operator that binds both the subject and the pronoun. The other way to resolve the anaphora is by means of covaluation as defined in (37). Reinhart uses this term rather than *coreference*, and defines it as an interface strategy available regardless the referential status of the antecedent (so it shows up also in quantified contexts, see example (39)), and it is not governed by considerations of the computational system. The pronoun *she* gets a value from the discourse storage, and this value can be *Lucie* (among others).

¹³ Note that Chomsky (1981) uses a crucially different definition of binding, which resorts to relations between arguments by means of coindexation (relations between identity variables). This enabled the formulation of the syntactic conditions on binding. I will come back to this issue at the end of this section.

(37) Covaluation: (Reinhart 2006:172)

α and β are covalued iff neither A-binds the other and they are assigned the same value.

Note that the pronoun *she* can also get the value *Lili*. This is called *covaluation*, and gives a similar interpretation (35d) as binding (35b). Nonetheless, the two construals (binding and covaluation) are different in their truth conditions. Sentences with pronouns like (33b) are ambiguous between a binding interpretation (35b) and a covaluation interpretation (35c). This ambiguity is easily seen in cases of ellipsis (38) and with the disambiguation element *only* in (39):

(38) a. Lili thinks she has gotten the flu, and Max does too. (Reinhart 2006:167)

b. BINDING: Lili (λx (x thinks x has gotten the flu) & Max (λy (y thinks y has gotten the flu))

c. COVALUATION: Lili (λx (x thinks z has gotten the flu) & Max (λy (y thinks z has gotten the flu) & $\underline{z=Lili}$)

The binding construal (38b) of the sentence (38a) yields the *sloppy reading*: Lili thinks that she (Lili) has the flu, and Max thinks that he (Max) has the flu too. On the other hand, the covaluation construal (38c) of (38a) gives the *strict reading*: Lili thinks that she (Lili) has the flu, and Max thinks too that she (Lili) has the flu (but does not mean that Max thinks that he has the flu). We see that the truth conditions of both construals are different: in the sloppy reading, both Lili and Max are thought to have the flu, whereas in the strict reading only Lili is thought to have the flu.

(39) a. Only Lucie respects her husband. (Reinhart 2006:168)

b. BINDING: Only Lucie (λx (x respects x's husband))

c. COVALUATION: Only Lucie (λx (x respects her husband) & $\underline{her = Lucie}$)

Reinhart (2000) argues that the element *only* is a disambiguator in sentence (39a). (39b) entails that unlike Lucie, other women do not respect their husbands. (39c) entails that other women do not respect Lucie's husband. The two construals are, again, truth-conditionally distinct.

Chomsky's (1981) notion of binding (see (41) in section 3.3.1.3), which is defined as a relation between variables (indexes), cannot account for the two construals of the

pronouns like (35b) and (35c) for *she* in (34b)¹⁴. Therefore, I will use Reinhart's definition of A(rgument)-binding in (36) to refer to binding in subsequent chapters. However, all the literature that will be reviewed in what rest of this section makes use of Chomsky's notion, which is based on indexes.

3.3.1.3. *Binding Theory*

There are certain pronominal elements that appear with reflexive verbs, and have a restricted syntactic distribution, i.e. there are restrictions on their occurrence that are imposed by the syntactic configuration.

- (40) a. $Lucie_i$ though that $she_{i/j}$ had the flu.
 b. $Lucie_i$ loves $herself_{i/*j} / her_{*i/j}$.
 c. $Lucie_i$ thought that $??/*herself_{i/*j}$ had the flu.

As seen in the previous section, the pronoun *she* in (40a) can refer to Lucie (due to binding or covaluation) or to another referent present in the discourse storage (valuation). Note however that the reflexive element *herself* in (40b) has to refer to *Lucie*. The pronoun *her* has to be used to refer to someone other than *Lucie*. Note that in that case, *her* cannot refer to *Lucie*, thus binding and covaluation of the pronoun is prohibited. We will see later that this is because pronouns like *her* cannot be locally bound, i.e. their antecedents have to be more distant (for example in a different clause). Moreover, *herself* shows syntactic restrictions as can be seen in (40c)¹⁵: its antecedent has to be local (a non-trivial notion to which I will devote the whole section 3.3.2), i.e. *Lucie* is not in the embedded clause but in the matrix clause, thus "too far" from *herself*.

Chomsky (1981) defines *binding* in (41) as a relation between variables (indexes), in order to account for the relation between arguments, i.e. pronouns and their antecedents. This definition enabled him to account for the syntactic restrictions found in the binding of different pronominal elements as seen in (40) above.

¹⁴ See the full discussion in Reinhart (2000, 2006). She further analyzes restrictions on these two anaphora resolution strategies, i.e. when binding can and must apply, and when covaluation applies as an interface strategy that requires reference-set computation.

¹⁵ Actually, *herself* in (40c) can be used, but only with an emphatic meaning. If that emphatic nuance does not suit the speech situation, *herself* is clearly ungrammatical.

(41) Binding: (Chomsky 1981:184)

α is X-bound by β if and only if α and β are coindexed, β c-commands α and β is in an X-position.

An element α binds another element β if they are coindexed, i.e. their identity variables are indistinguishable, and both elements are in a structural configuration of c-command, defined in (42).

(42) C-command: (Chomsky 1981:166)

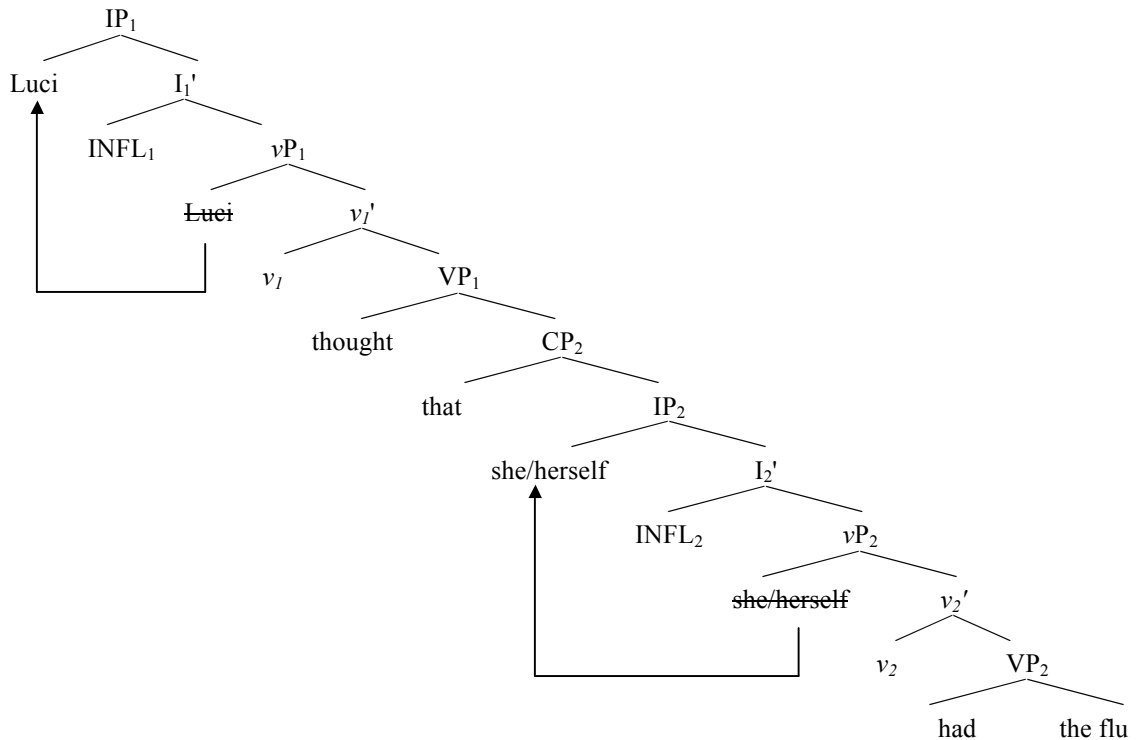
α c-commands β if and only if

- (i) α does not contain β
- (ii) Suppose that $\gamma_1, \dots, \gamma_n$ is the maximal sequence such that
 - a. $\gamma_n = \alpha$
 - b. $\gamma_i = \alpha^j$
 - c. γ_i immediately dominates γ_{i+1}

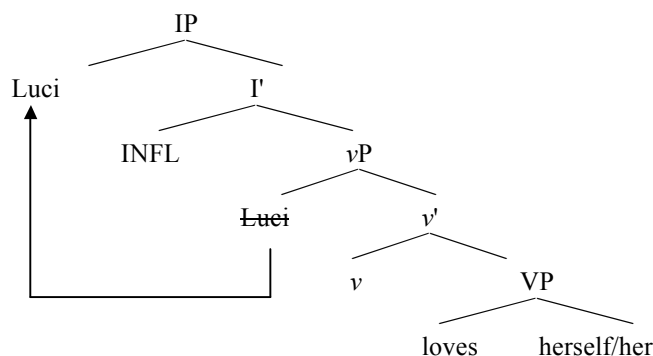
Then if δ dominates α , then either (I) δ dominates β , or (II) $\delta = \gamma_i$ and γ_i dominates β .

Let us come back to the examples in (40), repeated in (43) with a representation of their syntactic configurations:

(43) a. $Lucie_i$ thought that $she_{i/j} / ??/*herself_{i/*j}$ had the flu.



b. $Lucie_i$ loves $herself_{i/*j} / her_{*i/j}$.



In (43a) both *she* and *herself* are in a c-command relation with *Lucie*, their antecedent: the minimal maximal category dominating *Lucie* (CP_1) dominates *she* and *herself* too. Also in (43b) *herself* and *her* are c-commanded by their antecedent: the minimal maximal category dominating *Lucie* (CP) dominates *herself* and *her*. The difference between (43a) and (43b) is that the pronouns are in an embedded sentence in the first case and in the same clause in the second case. In other words, there is a CP node (CP_2) in (43a) that intervenes between the pronouns *she* and *herself* and their antecedent *Lucie* whereas there is no CP node in (43b) between the pronouns and their antecedent. We see then that the presence of intervening CP nodes is crucial for the syntactic formulation of the notion of locality.

Chomsky (1981) formulated the syntactic conditions on binding in (44), and distinguished two kinds of anaphora (pronouns): pronominals like *she* in (43a) and anaphors like *herself* in (43b). R-expressions are expressions that are able to get a referent directly from the context (exophoras).

(44) Binding Theory:

(Chomsky 1981:188)

- A. An anaphor is bound in its governing category.
- B. A pronominal is free in its governing category.
- C. An R-expression is always free.

R-expressions are "full nominals" like *Lucie* in (43). Condition C states that these expressions are never bound by another element because they have independent reference, i.e. they are exophoras as said before. Chomsky (1982), among others, stated that condition C was not part of the grammar, and reduced Binding Theory to conditions A and B. These conditions make use of the term *governing category* as a formalization of locality. Basically, what they say is that pronominals must be free (not

bound by any antecedent) in its local domain, whereas anaphors must be locally bound by an antecedent. We saw in (43) that the presence of a CP node is crucial for the local domain for anaphors and pronominals. Chomsky worked this out and formulated locality in the notion of *governing category* in (45). Other definitions have been proposed in order to accommodate the Binding Theory to data from languages other than English (see section 3.3.2).

(45) Governing category: (Chomsky 1981:220)

β is a governing category for α if and only if β is the minimal category containing α , a governor of α , and a SUBJECT (accessible to α)¹⁶.

(46) Government: (Chomsky 1981:163)

α governs β if and only if

(i) $\alpha = X^0$

(ii) α c-commands β and if γ c-commands β , then γ either c-commands α or is c-commanded by β .

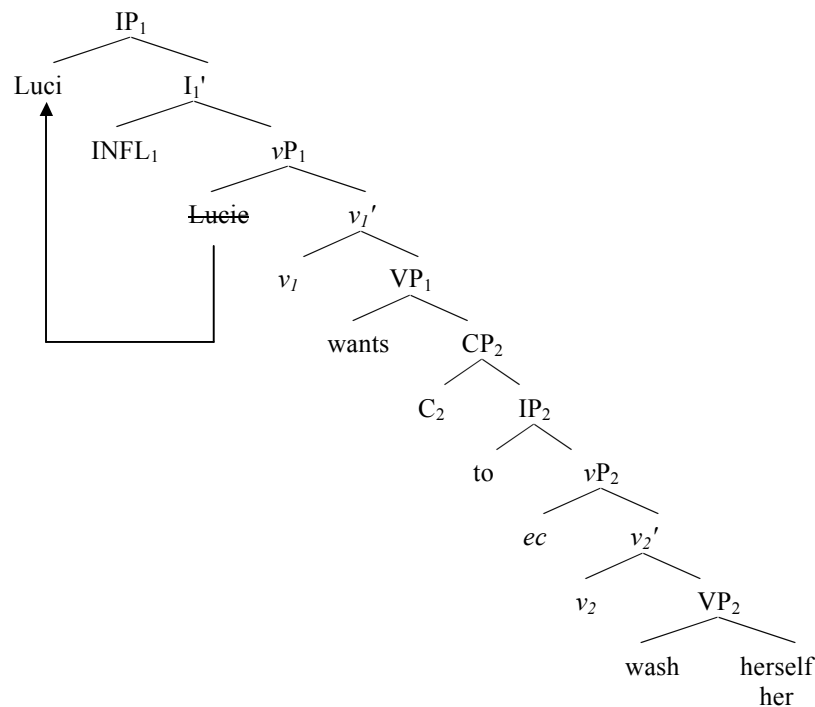
The governing category of *she/herself* in (43a) is the embedded clause (CP₂) because CP₂ is the minimal category that contains *she/herself*, a governor of them (INFL) and an accessible SUBJECT (Agr on INFL). Hence, *herself* is ruled out because it violates condition A: it is bound by *Lucie*, which is out of its governing category CP₂. The pronominal *she* is ruled in because it has to be bound outside its governing category, and *Lucie* is beyond CP₂.

In (43b) the governing category of *herself/her* is the IP because it is the minimal category that contains *herself/her*, a governor of them (the verb *love*) and an accessible SUBJECT (Agr on INFL and *Lucie*). *Her* is ruled out by condition B: as a pronominal, it cannot be bound by *Lucie* since it is within its governing category. *Herself* is ruled in because it is an anaphor, and has to be bound by an antecedent in its governing category, i.e. *Lucie* in IP.

The definition of *governing category* in (45) makes the correct predictions with regard the distribution of anaphors and pronominals within infinitival embedded sentences as in (47).

¹⁶ A SUBJECT is a subject and Agr in finite INFL.

(47) $Lucie_i$ wants to wash $herself_{i/*j}$ / *her.



The governing category of *herself/her* in (47) is not CP_2 but IP_1 . CP_2 contains *herself/her* and a governor (*wash*), but it does not contain an accessible SUBJECT: there is no subject and Agr of a non-finite INFL does not count as SUBJECT. Therefore, IP_1 is the minimal category that contains *herself/her*, a governor (*wash*) and an accessible SUBJECT (*Luci* and Agr in finite INFL). Then, the pronoun is ruled out by condition B and the anaphor is ruled in by condition A: it is bound by its antecedent *Lucie*, which is within its governing category IP_1 .

According to Reinhart (2000), Chomsky's definition of binding is not able to account for the two construals of the anaphora in (33b), and therefore I chose for Reinhart's definition of binding in (36). Furthermore, Reuland (2001) points out that the binding conditions of Chomsky's in (44) use both syntactic and semantic notions: *governing category* and *SUBJECT* are syntactic concepts, whereas *indices* and *binding* are semantic ones.

The Minimalist Program introduced by Chomsky (1995) aims to define the boundaries of the syntax in a principled way. The computational system of human language (C_{HL}) reflects the combinatorial properties of a purely morphosyntactic vocabulary. As C_{HL} is the optimal solution to map form and interpretation and so to meet the conditions imposed by the systems of thought and perception/articulation, a "perfect language" must obey the Inclusiveness Condition, which states that a structure

formed by the computation is constituted of elements already present in the lexical items selected. Thus no new objects can be added in the course of the derivation. As a consequence of the Inclusiveness Condition, indices, which are indispensable for the notion of binding, are not available within C_{HL} and this leads Chomsky (1995, 2001) to propose that binding conditions can only apply at the conceptual-intentional (C-I) interface (instead of at S-Structure like in the Government and Binding framework). This is another reason to prefer Reinhart's definition of binding rather than Chomsky's. Even so, there seem to be syntactic restrictions on the distribution of pronominals and anaphors as shown in (40). Reuland (2001) discusses other cases of binding (SE-anaphors) where there seems to be some syntactic residue too (what he calls the *Residual Condition A*). I will come back to this issue in section 3.3.3, where I will also mention some other problems.

3.3.1.4. *Summary*

In this section I have defined two ways of interpreting nominal items: deixis and anaphora. I have explained the main strategies of anaphora resolution (binding and valuation). Finally, I have introduced the Binding Theory, which is devoted to accounting for the referential interpretation of different types of nominals: R-expressions, pronominals and anaphors.

3.3.2. **Anaphors and binding domains**

We have seen in section 3.3.1.2 that Chomsky's (1981) Binding Theory has empirical problems with regard to the different anaphora construals of pronominals, as well as theoretical problems (derived from the Inclusiveness Condition) in section 3.3.1.3. Apart from these problems, others have been identified for the Binding Theory as originally stated in (41) and (44). Consider (48):

- (48) a. I_i want [$PRO_{i/*j/*arb}$ to read the Silmarillion].
 b. $John_i$'s friend $_j$ thinks that [$PRO_{i/j/k/arb}$ being naive is something bad].

These sentences are thought to have a non-phonetically realized anaphoric element (PRO, I will come back to this issue right below) in subject position that can be either

locally (48a) or non-locally (48b) bound¹⁷. This anaphoric element (whether pronominal or anaphor) seems to escape the binding conditions (44) proposed by Chomsky.

Furthermore, the Binding Theory predicts a perfect complementary distribution between anaphors and pronominals. This is nonetheless contradicted by data from many languages. Examples (49), (50), and (51) from Reinhart & Reuland (1993) show that condition A can be violated, and that pronominals and anaphors need not always be in complementary distribution (see section 3.3.2.3). On the other hand, Raposo's (1985) example in (52) shows that also condition B can be violated in European Portuguese under certain conditions (see section 3.3.2.2).

(49) Max_i said that the queen invited both Lucie and [himself_i/him_i] for tea. (R&R 1993:675)

(50) The queen invited both Max and [myself/me] for tea. (R&R 1993:675)

(51) Lucie_i saw [_{NP} a picture of her_i / herself_i] (R&R 1993:661)

(52) O Manel_i deseja que ele_{*i/j}/pro_{*i/j} lê mais livros. (Raposo 1985:77)
 The Manel wishes that him / pro read more books
"Manel wishes that he reads more books."

In this section, I will review the problems that the canonical Binding Theory has had as it concerns the notion of locality, which will be defined by the term *binding domain* (which is, in its turn, an extension of the concept *governing category*)¹⁸. We will see that the binding domain for pronouns can be extended in certain cases, and that there are different binding domains for anaphors both across and within languages, although this variation is not free and can be explained in a principled way. Nonetheless, this requires going one step further from the canonical Binding Conditions proposed by Chomsky, and distinguish at least two kinds of anaphors: SE-anaphors vs. SELF-anaphors (following Reinhart & Reuland 1993), besides the special case of the pronominal anaphor PRO as well as the logophoric uses of the anaphors.

¹⁷ In (48b) PRO can be non-locally bound in the sense that a nominal item can bind it even if there is no c-command relation between them both, as the case of *John* and PRO in (48).

¹⁸ The theoretical problems the Binding Theory has to cope with, such as those derived from the notion of *government* and by the Inclusiveness Condition, will be discussed in section 3.3.3.

3.3.2.1. *Beyond the Binding Theory: Control and PRO*

In section 3.3.1.3 we saw that the Binding Theory defines two kinds of pronouns: pronominals like *she* in (40a) anaphors like *herself* in (40b). It also predicts a perfect complementary distribution of them both. Nonetheless, consider the sentences in (48) above.

Unlike raising constructions in (53), control constructions in (48) are not derived by movement due to theta considerations. More concretely, movement to theta positions cannot be resorted to since that would lead to a violation of the Theta Criterion, which states that a nominal cannot bear two theta roles at a time (though see Hornstein 1999 and discussion in section 3.2.2.1 above).

(53) John_i appears [t_i to have read the Silmarillion].

Due to the Extended Projection Principle (EPP: all sentences must have a syntactic subject) and since sentences in (48) both have an implicit subject, it has been proposed that an empty category (called PRO by Chomsky 1982) occupies the subject position of the embedded clause.

Note, however, that PRO has special properties with regard to the Binding Theory, as said before: it seems to be subject both to condition A (48a) and condition B (48b). Hence, PRO falls out of the (standard) Binding Theory. Chomsky (1982) proposes that its syntactic distribution is accounted for by the PRO theorem in (4), repeated in (54), which is derived from the observation that PRO is subject to both conditions A and B of the Binding Theory. The only way to resolve such a contradiction is that PRO has no governing category, and this is achieved if PRO is ungoverned. This happens if PRO is in subject position of tenseless sentences: there is no category that contains PRO, a SUBJECT accessible for PRO and a governor of PRO since in that position, nothing properly governs PRO (non-finite INFL is not a proper governor). Hence, an independent module of the grammar, different from the Binding Theory, had to be proposed to account for the syntactic distribution and semantic interpretation of PRO: Control Theory.

(54) PRO Theorem: *(Chomsky 1981:191)*
PRO must be ungoverned.

However, Manzini (1983) proposed that Control Theory is constructed on essentially the same notions on which Binding Theory is constructed in Chomsky (1981). She

defended that PRO is a pure anaphor that has to be ungoverned due to Case considerations rather than binding requirements. In other words, PRO is an anaphor that is not spelled-out. Therefore it has to be Caseless, and hence the requirement of PRO being ungoverned. Besides the notion of governing category in (45) above, Manzini defined the notion of *domain-governing category* in (55), and argued that this is the notion of locality that holds for anaphors without governing category like PRO, i.e. that an anaphor without a governing category must be bound in its domain-governing category.

- (55) γ is a domain-governing category for α iff: (Manzini 1983:433)
- a. γ is a governing category for the c-domain of α , and
 - b. γ contains a subject accessible to α .

Manzini revised Condition A as in (56a) so as to capture the distribution of anaphors (including PRO), and gave a revised version of the Binding Theory in (56).

- (56) Binding Theory (revisited): (Manzini 1983:432)
- A. An anaphor is bound in its governing category and its domain-governing category.
 - B. A pronominal is free in its governing category.

PRO is subject to the revisited condition A (56a), and hence falls within the modified Binding Theory (56). Its special distribution is due to its lack of structural Case.

In conclusion, Manzini defined PRO as a pure anaphor whose special distribution is due to its Case properties. It falls under the scope of the Binding Condition once the locality notion that holds for anaphors is redefined as the domain-governing category rather than the governing category.

As a conclusion, we see that notion of locality, along with Case specification, is the central point that accounts for the distribution of PRO according to Manzini¹⁹. Therefore, Control Theory need not be stipulated as an independent module of the

¹⁹ There are many authors that, like Manzini (1983), have tried to reduce Control Theory to Binding Theory, Movement Theory or some other mechanism already present in the theoretical machinery. These proposals have been reviewed in section 3.2. In section 3.4 I will propose a new analysis of obligatory and non-obligatory control that, following the spirit of Manzini's proposal, reduces control to binding as conceived by Reuland (2001).

grammar. The hypothesis put forward in section 3.4 will lead us to reach the same conclusion as this one from the work of Manzini's.

3.3.2.2. *Locality and condition B: Disjoint Reference Effect and Short Distance Pronouns*

The notion of locality has been a problem for Binding Theory. We have seen in the previous section that Manzini proposed to change the notion of locality from governing category to domain-governing category so that the Binding Theory could account for PRO (and control phenomena) too. In this section we will see that Romance languages pose some other problems related to the notion of locality that applies to binding of pronominals.

Raposo (1985) identified the phenomenon of *disjoint reference effect* (DRE hereinafter). Let us look at (57)²⁰:

- (57) a. O Manel_i pensa que ele_i/pro_i lê bastantes livros. (Raposo 1985:77)
 The Manel thinks that he / pro reads many books
 "Manel_i thinks that he_i reads many books."
- b. O Manel_i deseja que ele_{*i/j}/pro_{*i/j} lê mais livros. (Raposo 1985:77)
 The Manel wishes that he / pro reads more books
 "Manel_i wishes that he_{*i/j} reads more books."

Verbs like *pensar* (think) in (57a) are E(pistemic)-verbs that select tensed clauses, usually marked by indicative morphology. The pronominal subject of the embedded clauses that are complement to E-verbs can corefer with the verb in the matrix sentence. This is predicted by condition B: the governing category of *ele* (he) is the tensed embedded sentence and there, *ele* must be free though it can be bound by an antecedent out of the embedded sentence.

Desejar ("wish") in (57b) is a W-predicate that selects non-tensed clauses in the sense that the tense-frame of these embedded clauses are somehow dependent on the tense of the matrix clause, usually referring to some sort of "future orientation". This is usually signalled by the use of subjunctive mood. The pronominal subject of an embedded clause complement to a W-predicate cannot be bound by the matrix subject. Raposo called this the *disjoint reference effect* (DRE hereinafter)²¹. This is not predicted

²⁰ Examples (57) and (59)-(64) are from Raposo (1985).

²¹ This effect has also been called *obviation*. See Comorovski (1985) and Luján (1999).

by the Binding Theory: the governing category of *ele* in (57b) is the embedded clause as in (57a). However, *ele* must be free in the matrix sentence too. It seems that the governing category, or rather the binding domain (the domain where the pronominal must be free), extends beyond the embedded clause as it happens with infinitivals (which is predicted by Binding Theory since the governing category of an infinitival is its matrix clause). The DRE is observed in Spanish too, as can be seen in (58).

- (58) a. Juan_i piensa que él_{i/j}/pro_{i/j} lee bastantes libros.
 Juan thinks that he / pro reads many books
"Juan_i thinks that he_{i/j} reads many books."
- b. Juan_i desea que él_{*i/j}/pro_{*i/j} lea más libros.
 Juan wishes that he / pro reads more books
*"Juan_i wishes that he_{*i/j} reads many books."*

Raposo proposes that the differences observed between E-predicates and W-predicates are due to the presence of a verbal operator [TENSE] on the node Comp (which is not to be confused with the [Tense] feature on INFL). This verbal operator defines whether the clause is semantically autonomous or dependent with regard to its time frame specification. W-predicates select [-TENSE] clauses (usually marked with subjunctive mood) whereas E-predicates select [+TENSE] clauses (usually marked with indicative mood).

The DRE takes place because in Romance, [+TENSE] on Comp rather than Agr on INFL, creates an opaque domain for a pronoun in subject position. In other words, Raposo states that [+TENSE] is a SUBJECT rather than Agr for pronominals in Romance. Not only the fact of being complement of W-predicates does modify the binding domains. Auxiliaries, which are thought of as tense operators, allow the embedded clauses of W-predicates to have a time-independent frame as in (59). Hence, auxiliaries like *ter* (have) weaken the DRE of embedded clauses of W-predicates as in (60) versus (61), and so do modals like *poder* (can) in (62) versus (64) and "aspectual" auxiliaries like passive *ser* (be) in (63) versus (64).

- (59) a. Eu receio que a Maria tenha encontrao o Antonio. (Raposo 1985:85)
"I fear that Maria has met Antonio."
- b. Eu desejo/quero que a Maria tenha ganho a prémio. (Raposo 1985:85)
"I wish/want that Maria has won the prize."

- (60) a. O António receia que (ele) tenha bebido a água envenenada. (Raposo 1985:86)
"Antonio fears that (he) has drunk the poisoned water."
 b. A Maria preferia que (ela) não tivesse encontrado o Manel. (Raposo 1985:86)
"Maria preferred that (she) had not met Manel."
- (61) a. O António receia que (ele) beba a água envenenada. (Raposo 1985:86)
"Antonio fears that (he) drinks the poisoned water."
 b. A Maria preferia que (ela) não encontrasse o Manel. (Raposo 1985:86)
"Maria preferred that (se) did not met Manel."
- (62) a. O Manel exige que (ele) possa ver o seu advogado. (Raposo 1985:86)
"Manel requires that (he) be able to see his lawyer."
 b. O manel recomendou que (ele) pudesse escolher a equipa. (Raposo 1985:86)
"Manel recommended that (he) might choose the team."
- (63) a. O Manuel deseja que (ele) seja admitido no concurso. (Raposo 1985:86)
"Manel wishes that (he) be admitted in the contest."
 b. O Manel recomendou que (ele) fosse escolhido para a equipa. (Raposo 1985:86)
"Manel recommended that (he) be chosen for the team."
- (64) a. O Manuel exige que (ele) veja o seu advogado. (Raposo 1985:87)
"Manel requires that (he) sees his lawyer."
 b. O Manel recomendou que (ele) escolhesse a equipa. (Raposo 1985:87)
"Manel recommended that (he) chose the team."

Raposo concludes that all these elements ([±TENSE], auxiliaries and modals) are verbal operators that introduce opacity for pronominals and condition B. They take the clause as their operand and create a local opaque domain where pronominals are free.

Nevertheless, anaphors cannot appear in subject position of W-embedded clauses as can be seen in (65).

- (65) *O Manel deseja que si próprio ganhe o prémio. (Raposo 1985:88)
"Manel wishes that himself win the prize."

Based on these data, Raposo reaches the conclusion that Agr does not create opacity for pronominals unlike verbal operators. In other words, in Romance Agr is not a SUBJECT. On the other hand, Agr does create opacity for anaphors so Agr is a SUBJECT for anaphors.

The conclusion is that in Romance, the governing categories are different for pronouns and anaphors. Given the definition (45) of governing categories, it is the SUBJECT that varies in Romance for pronouns and anaphors. The SUBJECT for pronouns is a verbal operator whereas it is Agr for anaphors.

Apart from the DRE in Romance, condition B may be affected by other factors in other languages. Pronouns can be locally bound in directional or locational PPs (66) as well as in complex or picture NPs in (67), as noted by many authors like Manzini (1983), Reinhart & Reuland (1993) and Tenny (2004). In these syntactic configurations, the complementary distribution of pronominals and anaphors predicted by the Binding Theory is borne out.

(66) Directional or locational PPs: (Tenny 2004:1)

a. John_i has no covering over him_i.

(67) Representational NPs:

a. Lucie_i saw [_{NP} a picture of her_i / herself_i]. (R&R 1993:661)

b. Max_i likes [_{NP} jokes about him_i / himself_i]. (R&R 1993:661)

Tenny calls these instances of locally bound pronominals *short distance pronouns*, and shows that their acceptability is closely related to discourse factors, like *long distance anaphora* (see section 3.3.2.3 below), which Tenny develops in a Grammar of Sentience (Tenny 2004).

To conclude, we see that condition B does not always hold for pronominals as originally stated by Chomsky (1981). According to Reuland & Koster (1991), this condition seems to be quite robust compared to Condition A. Much more variation is observed in the binding domains of anaphors (the domains where anaphors must be bound), a point to which we turn on in the next section.

3.3.2.3. *Locality and condition A: Long distance anaphora*

Condition A of the Binding Theory does not hold for anaphors as stated in Chomsky (1981) in many languages, as it happens with condition B in Romance W-predicates, directional PPs and representational NPs. The phenomenon of non-locally bound anaphors is most commonly referred to as *long distance anaphora*, i.e. anaphors that can be bound beyond their binding domain and violate thus condition A, like the Icelandic examples in (68) from Reuland (2006b).

- (68) a. Jón_j skipaði Pétur_i [að PRO_i raka_{infinitive} sig_{i,j,*k} á hverjum degi].
 (Reuland 2006b:544)
 John ordered Peter [to PRO shave sig every day]
- b. Jón_j segir [að Pétur_i raki_{subjunctive} sig_{i,j,*k} á hverjum degi].
 (Reuland 2006b:544)
 John says [that Peter shaves sig every day]
- c. Jón_j veit [að Pétur_i rakar_{indicative} sig_{i,*j,*k} á hverjum degi].
 (Reuland 2006b:544)
 John knows [that Peter shaves sig every day]

The phenomenon of long distance anaphora has led to a modification of the Binding Theory by adjusting the definition of the binding domains for anaphors across languages and within particular languages. The binding domain is defined by the governing category, which can be modified by introducing an opacity factor F.

- (69) Governing category (revisited): (Reuland & Koster 1991:2)

β is a governing category for α iff β is the minimal category containing α , a governor of α , and F (F an opacity factor).

Long distance anaphora is thus defined as anaphors whose opacity factor F is beyond their SUBJECT.

In principle, the factor F could be freely defined so that unlimited variation in the binding domains for anaphors would be possible. However, Reuland & Koster's (1991) survey of languages shows that F for anaphors does not vary freely and hence principled restrictions of F can be defined. Reuland & Koster (1991:10-11) give the definition in (70a) for long distance anaphor (LDA) and their properties in (70b-e).

- (70) a. LD-anaphors allow an antecedent outside the governing category as defined in (45).
- b. The antecedents of LD-anaphors are subject to a more restrictive prominence condition than c-command. The most common requirement is that the antecedent must be a subject.
- c. LDA is restricted to reflexives. Reciprocals are not allowed as LDAs.
- d. LDAs are morphologically simplex. Morphologically complex anaphors are local.
- e. Outside the local domain there is no complementarity between pronouns and anaphors.

Building on the empirical evidence from their survey of languages, Reuland & Koster define three binding domains for anaphors:

- (71) Binding domains: (Reuland & Koster 1991:23-24)
- a. DOMAIN 1: the governing category when F = accessible subject.
 - b. DOMAIN 2: the governing category when F = finite INFL (subjunctive or indicative).
 - c. DOMAIN 3: the governing category when there is no F.

Long distance anaphors are usually morphologically simpler than local anaphors (70d), as shown in the contrast of short and long distance anaphors in (72). Reinhart & Reuland (1993) called them SE-anaphors so as to distinguish them from the morphologically complex anaphors (SELF-anaphors) as that in (72a). They observed that only SE-anaphors could be non-locally bound in domain 2 or beyond this domain (which always implies discourse factors or logophoricity). On the other hand, only SELF-anaphors as (72a) but not SE-anaphors as (72b) can license reflexive readings of verbs that have not undergone any lexical operation on their thematic grid²². This issue will be crucial later in chapters 5 and 6 when addressing the issues of reflexivity and unaccusativity.

- (72) a. Harry loves **himself**.
- b. Jón_j veit [að Péturi rakar_{indicative} **sig**_{i,*j,*k} á hverjum degi].
- (Reuland 2006b:544)
- John knows [that Peter shaves **sig** every day]

To sum up, Reuland & Koster defined actually two binding domains: the *local binding domain*, which is the governing category and the subject as the opacity factor. This is the binding domain where Chomsky's (1981) binding conditions hold, and where complementary distribution between pronominals and anaphors is found. The other domain is the *extended binding domain*, which corresponds with the first finite INFL (whether indicative or subjunctive). No binding domain can be defined beyond the extended domain.

²² Note that in (72a) the verb "love" is not marked as reflexive verb in the lexicon and therefore it requires a morphologically complex anaphor in order to trigger a reflexive reading, whereas the verb "shave" in (72b) is marked as reflexive verb in the lexicon and hence it can use a long distance anaphor. We will come back to this different classes of verbs in chapter 5.

Based on these binding domains for anaphors, three sorts of binding can be defined (Reuland & Koster 1991:23-24):

- LOCAL BINDING: it takes place between an anaphor and its antecedent in the local binding domain. There is alternation between pronominal and anaphors (though see picture NPs in Reinhart & Reuland 1993, Tenny 2004, among others), i.e. the (canonical) Binding Theory holds. SE-anaphors, which are long distance anaphors, can be locally bound when the thematic grid of the verb has been somehow modified (Reuland & Koster 1991).
- MEDIUM DISTANCE BINDING: it takes place between a long distance anaphor (a SE-anaphor) and its antecedent in the extended binding domain. SELF-anaphors cannot be bound in the extended domain but need to be locally bound.
- "REAL" LONG DISTANCE BINDING: it takes place between a long distance anaphor and its antecedent beyond the extended domain. It always involves logophoricity, which is a form of deixis that refers to event orientation (for a more extended discussion on logophoricity see Reuland 2006b, 2006c). Residual cases of logophoric interpretation of SELF-anaphors can be found in English, as will be shown in chapter 5 section 4.2.1.

Finally, according to the binding domains of anaphors (whether local or non-local), their morphological realization (simplex or complex) and whether they license reflexive readings with verbs that are not marked as reflexive verbs in the lexicon, we can distinguish two kinds of anaphors²³ (Reinhart & Reuland 1993):

- a. **SELF-anaphors:** are morphologically complex, they license reflexive readings with verbs that are not specified as reflexive in the lexicon and they must be locally bound. They are anaphors like English *myself*, *herself* and *himself* in (72a).
- b. **SE-anaphors:** are morphologically simplex, they cannot license reflexive readings with verbs that are not specified as reflexive in the lexicon and they can be locally bound, as Dutch *zich* in (73), but do not need to, i.e. they can be non-locally bound as Icelandic *sig* in (68).

(73) Jan scheerde **zich**.

Jan shaved *zich*

"Jan has shaved (*himself*)."

²³ I will come back and present a deeper description of the nature, feature composition, similarities and differences between SE-anaphors and SELF-anaphors in chapter 5.

3.3.2.4. Summary

To conclude this section, we see that the canonical Binding Theory cannot be maintained as originally proposed by Chomsky (1981). It faces serious theoretical problems (violation of the Inclusiveness Condition, mixture of semantic and syntactic notions) as well as many empirical shortcomings, most of them reducible to the notion of locality that applies to binding (control phenomena and PRO, DRE, short distance pronouns, long distance anaphora).

We have seen that two types of anaphors can be defined: SELF-anaphors or morphologically complex anaphors always locally bound and able to license reflexive readings, and SE-anaphors that can be either locally bound (if a operation on the thematic grid of the verb has taken place) or non-locally bound (long distance anaphora; medium distance binding and logophoricity).

Besides these two types of anaphors (SE- vs. SELF-anaphors), PRO is a pronominal anaphor that escapes the Binding Theory in such a way that another module of the Grammar has been defined (Control Theory). I will show in section 3.4.1 that PRO has in common with SE-anaphors that it can be non-locally bound and does not license reflexive readings. So will I show that PRO has in common with SELF-anaphors that can be locally bound without any modification of the thematic grid of the verb. Finally, I will show that PRO has in common with pronominals that it can get a referent from the discourse.

3.3.3. Anaphors, binding and syntactic chains

Reuland (2001) points out that the canonical binding conditions (Chomsky 1981) in (74) use both syntactic and semantic factors: notions such as *governing category* and *subject* are syntactic whereas *indices* and *binding* are semantic concepts.

(74) Binding Canonical Conditions: (Chomsky 1981:188,211)

A. An anaphor is bound in its governing category.

B. A pronominal is free in its governing category

where β is a governing category for α if and only if β is the minimal category containing α , a governor of α , and a SUBJECT (accessible to α); and α binds β if and only if α and β are coindexed and α c-commands β .

The MP aims to define the boundaries of the syntax in a principled way. The computational system of human language (C_{HL}) reflects the combinatorial properties of

a purely morphosyntactic vocabulary. As C_{HL} is the optimal solution to map form and interpretation and so to meet the conditions imposed by the systems of thought and perception/articulation, a "perfect language" must obey the Inclusiveness Condition, which states that a structure formed by the computation is constituted of elements already present in the lexical items selected. Thus no new objects can be added in the course of the derivation. As a consequence of the Inclusiveness Condition, indices, which are indispensable for the notion of binding, are not available within C_{HL} and this leads Chomsky (1995, 2001) to propose that binding conditions can only apply at the conceptual-intentional (C-I) interface²⁴.

Reuland argues that despite the fact that binding takes place at the C-I interface (see the logical-syntax based definition of A-binding in (75)), there is indeed an irreducible syntactic residue in binding, a residual Condition A. This is to say that certain binding dependencies (namely, those between a DP and a SE-anaphor, such as Dutch *zich*) are encoded by C_{HL} by means of syntactic chains. Reuland (2001) uses feature checking and chain composition as in Chomsky (1995). The operation *Agree* did not exist yet in the system (Move-F were used instead). Nonetheless, as noted by Reuland, it is expected that the system is compatible with *Agree* though no implementation is given until Teomiro (2005) and Reuland (2006a). The implementation I will develop below follows my previous work (Teomiro 2005) and it is based upon chains formed by feature sharing (Pesetsky & Torrego 2007).

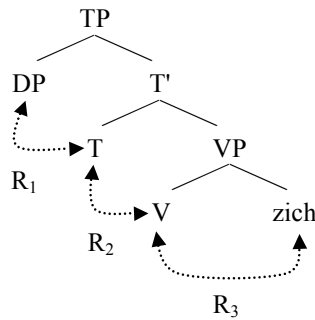
(75) A-binding (logical-syntax based definition): (Reinhart 2006:166)
 α A-binds β iff α is the sister of a λ -predicate whose operator binds β

Informally stated, a DP in subject position enters into a relation with the I system (represented as R_1 in (76) below) as well as a DP object enters into a relation with the V system (R_2). The I and the V systems enter too into a relation (R_3). Composition of $R_1+R_2+R_3$ into a syntactic chain takes place provided the object is not fully specified for φ -features. If so, the interpretive system interprets the chain as a binding relation. SE-anaphors are not fully specified for φ -features, e.g. Dutch *zich* is specified for the feature person but not for gender and number. Hence, chain formation will relate *zich* in object position to a suitable DP in subject position, and at the C-I interface the DP will be interpreted as the binder of *zich*. In cases where the SE-anaphor is within a PP, the

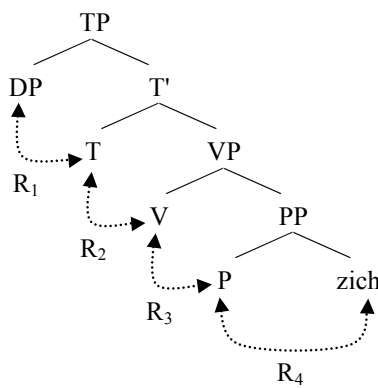
²⁴ See Lebaux (2009) for other proposals about the application of Binding Theory.

properties of P^{25} will determine whether composition is possible or not, i.e. whether R_3 in (76b) is allowed or not.

(76) a.



b.



The binding relation between two nominals can be thus encoded in two ways: directly at the C-I interface by A-binding (in the case of pronominals) or by means of a syntactic chain coded by C_{HL} that is translated to A-Binding at C-I (SE-anaphors). The latter is affected by syntactic restrictions, such as c-command or locality, that hold for any other syntactic chain. This is what Reuland calls *residual condition A*.

Pronominals can also be assigned a value from discourse storage without being bound. This is covaluation at discourse level, and can be seen in (77), where the pronominal is either interpreted as a bound variable (the sloppy reading) or assigned a value from the discourse storage (the strict reading). This indicates that at the C-I interface, pronouns can be translated either as expressions receiving a value directly from a discourse storage or as variables to be bound by an antecedent.

- (77) a. Bill liked his cat and Charles did too (Reuland 2001:447)
 b. Bill λx (x liked a's cat) & Charles λx (x liked a's cat)
 c. Bill λx (x liked x's cat) & Charles λx (x liked x's cat)

²⁵ See footnote 15 in Reuland (2001).

There are thus three ways in which a pronoun (pronominals and SE-anaphors) can be interpreted, according to Reuland (2001:473):

- a) Syntactic chains translated to A-binding → by C_{HL} and at C-I
- b) Variable binding by means of A-binding → at C-I
- c) Coreference by accessing discourse storage → covaluation at discourse level

Economy holds (Reuland 2001:473-474): encoding a dependency by C_{HL} requires fewer interpretative steps than encoding the same dependency by A-binding at C-I, which in turn requires fewer interpretative steps than encoding that same dependency by accessing the discourse storage (Reuland 2001).

Finally, it should be noted that, following Bouchard (1984), much of the G&B literature takes it that anaphors in general, and SE-anaphors in particular, have to be bound in order to be interpreted. This is based on the assumption that in order for an argument α to be interpreted, it must have a full specification for φ -features. Reuland (2001) shows that this does not follow from any deep principle of the grammar and that it is more accurate to state that if an element α has fewer φ -features than β , there are fewer constraints on the interpretation of α than on the interpretation of β ²⁶. From this, it follows that there is no intrinsic property of anaphors that prohibits an unbound interpretation. Unbound SE-anaphors can be attested in Icelandic and Latin (Reuland 2006c and references found in there).

3.3.4. Agree-chains

As I said in the introductory chapter, I will follow Pesetsky & Torrego's (2007) conceptualization of feature valuation as feature sharing. As a result of the *Agree* operation, feature co-valuation takes place and an agree-chain is established (Reuland 2006a; Teomiro 2005).

In (78) below the notation for agree-chains in general is specified.

(78) Agree-chains notation:

- a. $\alpha [uR[-]] \& \beta [iR[val]] \rightarrow Agree \rightarrow \alpha [uR[val]] \& \beta [iR[val]]$
- b. $R \{ \alpha, \beta \}$

²⁶ This will be crucial in the distinction of PRO and *pro* and why PRO is incompatible with φ -features on T (see section 3.7.1.).

I will distinguish two types of agree-chains (for the time being) that are necessary to develop the analysis of control in the next section:

(79) Types of agree-chains (to be revisited in chapter 5):

- a. **\varnothing -chain:** agree-chain formed when two or more lexical items share one or more \varnothing -features, as *niña* (child) and *bonita* (nice) in (82):
- b. **Tns-chain:** agree-chain formed when two or more lexical items share a Tns (tense) feature as in (80) and (81).

Below, there are three examples of different agree-chains that can be formed in a syntactic derivation: (80) shows a Tns-chain between the verb, the head Ts and *Bilbo* in order to share the feature Tns-s (NOM). Another Tns-chain can be found in (81) where *the One Ring*, the verb and the head To share the feature Tns-o (ACC). Finally, in (82) a \varnothing -chain is represented where *niña* and *bonita* share the \varnothing -feature gender.

(80) a. Bilbo stole the One Ring²⁷

- b. $\alpha [uT[-]] \& \beta [iT[-]] \& \gamma [uT[NOM]] \rightarrow Agree \rightarrow \alpha [uT[NOM]] \& \beta [iT[NOM]] \& \gamma [uT[NOM]]$
- c. Bilbo $[uT[-]] \& Ts [iT[-]] \& stole [uT[NOM]] \rightarrow Agree \rightarrow Bilbo [uT[NOM]] \& Ts [iT[NOM]] \& stole [uT[NOM]]$
- d. Tns₁-chain {Bilbo, Ts, stole}

(81) a. Bilbo stole the One Ring

- b. $\alpha [uT[-]] \& \beta [iT[-]] \& \gamma [uT[ACC]] \rightarrow Agree \rightarrow \alpha [uT[ACC]] \& \beta [iT[ACC]] \& \gamma [uT[ACC]]$
- c. the One Ring $[uT[-]] \& To [iT[-]] \& stole [uT[ACC]] \rightarrow Agree \rightarrow the One Ring [uT[ACC]] \& To [iT[ACC]] \& stole [uT[ACC]]$
- d. Tns₂-chain {the One Ring, To, stole}

(82) a. $\alpha [u \varnothing [-]] \& \beta [i \varnothing [val]] \rightarrow Agree \rightarrow \alpha [u \varnothing [val]] \& \beta [i \varnothing [val]]$

- b. *niña* $[i\varnothing_{gender}[fem]] \& bonit_ [u\varnothing_{gender}[-]] \rightarrow Agree \rightarrow niña [i\varnothing_{gender}[fem]] \& bonita [u\varnothing_{gender}[fem]]$
- c. \varnothing -chain { *niña*, *bonita* }

²⁷ In this example, I follow Pesetsky & Torrego's (2007) idea that the value of the Tns-s feature (NOM) is introduced in the derivation by the verb. I argue that also the value of the Tns-o feature (ACC) is introduced in the derivation by the verb (despite the fact that both these features are interpretable in the Ts/To heads).

3.4. The null SE-anaphor PRO

The analysis of control I put forward in this section follows Landau (2004b) in analyzing PRO as a phonetically null SE-anaphor. However, I will dispose of the Calculus and the function of [R] feature in Landau's analysis²⁸ by relying on Reuland's (2001, 2008) version of the binding theory. In addition I will adopt the theory of structural Case and feature valuation developed by Pesetsky & Torrego (2004, 2007). These authors have shown that structural Case is uninterpretable T(ense) on DP (Pesetsky & Torrego 2004) and that features can be valued by feature sharing (Pesetsky & Torrego 2007). As I will show, these approaches taken together provide an accurate account of control phenomena and the distribution of OC & NOC PRO in a substantial domain. The questions why PRO is incompatible with overt ϕ -features on T (Agr), and why PRO is phonetically null, will also receive a natural explanation. The same holds true of the relation between PRO, T and structural Case.

3.4.1. The interpretation of PRO

The hypothesis I propose is that PRO is a phonological null SE-anaphor with interpretable yet *unvalued* ϕ -features (at least person and number) and an unvalued and uninterpretable T feature, i.e. structural Case. I assume PRO is [+R] since in NOC cases it can pick up a value directly from the context. Hence, I also assume that it has a grammatical number feature, which enables the SE-anaphor to be referentially independent and go beyond its basic deictic use (Reuland 2006c). In this respect, PRO and *zich* differ from one another in that only PRO can refer to an entity without resorting to binding or logophoricity. In table (83), a comparison of the feature composition of the Dutch SE-anaphor *zich* vs. the null SE-anaphor PRO can be seen.

²⁸ I will assume that PRO is [+R], as can be clearly seen in NOC contexts.

(83) Feature composition *zich* vs. PRO:

	zich	PRO
Referential independency	[-R]	[+R]
φ -Features	$i \varphi_{\text{person}} [3^{\text{rd}}]$	$i \varphi_{\text{person}} [-]$ $i \varphi_{\text{number}} [-]$
Structural Case	uT [-]	uT [-]
Phonological content	yes	no

I will assume that the φ -features of PRO need not be valued and erased in order for the derivation to converge (for a similar proposal see Holmberg 2005) since they are interpretable. However, it is worth noting that although these φ -features end up unvalued, they can nevertheless delete uninterpretable instances of φ -features since they themselves are interpretable.

The φ -features of PRO can get valued by forming a chain with a controller or with the φ -features of the closest T, yet this case will be put aside until the section 3.7.1, where *pro* is discussed. If the φ -features of PRO are not valued, nonetheless, it is as if they did not exist, this is to say that they have no effect whatsoever. The fewer φ -features a pronoun has, the freer its interpretation will be (Reuland 2001): (84a) shows that Dutch *zich* can only refer to DPs that, in their turn, refer neither to the speaker nor to the addressee, due to its defined feature person. PRO, on the contrary, may refer to any DP, may this be speaker, addressee, both of them or neither of them, because of its unvalued φ -features, as in (84b).

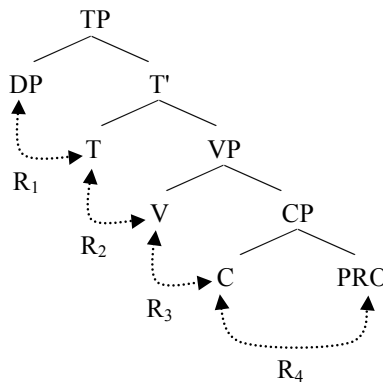
- (84) a. Hij/zij/*ik/*jij voel(t) zich niet goed
 He/she/*I/*you feels zich not well
 b. I/you/he/Ann/John want(s) PRO to read the book

PRO is interpreted depending on the structural configuration where it is located. When a syntactic chain can be formed, PRO will have OC characteristics. When this option is not present, PRO will behave as a pronominal, i.e. NOC will result.

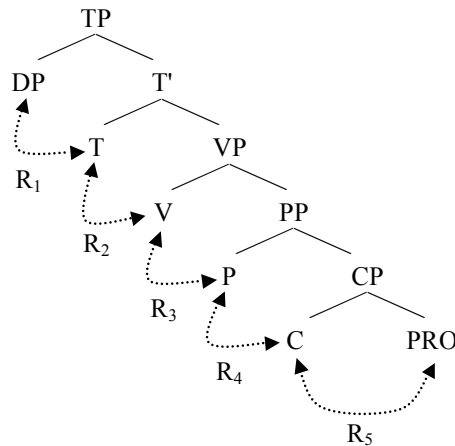
3.4.1.1. *Obligatory control*

When PRO can be bound by a DP by means of a syntactic chain, it behaves like a bound SE-anaphor and this yields OC PRO (its syntactic & semantic characteristics being derived from chain formation): we saw in section 3.3.3 that composition of the relations between a DP subject and a SE-anaphor object yields a dependency between a DP and a SE-anaphor which is interpreted as A-binding if and only if the SE-anaphor is φ -defective. The same picture emerges with PRO: since PRO has all its φ -features unvalued, the dependency between PRO and its controller may (and indeed must) be interpreted as binding at the C-I interface.

(85) a.



b.

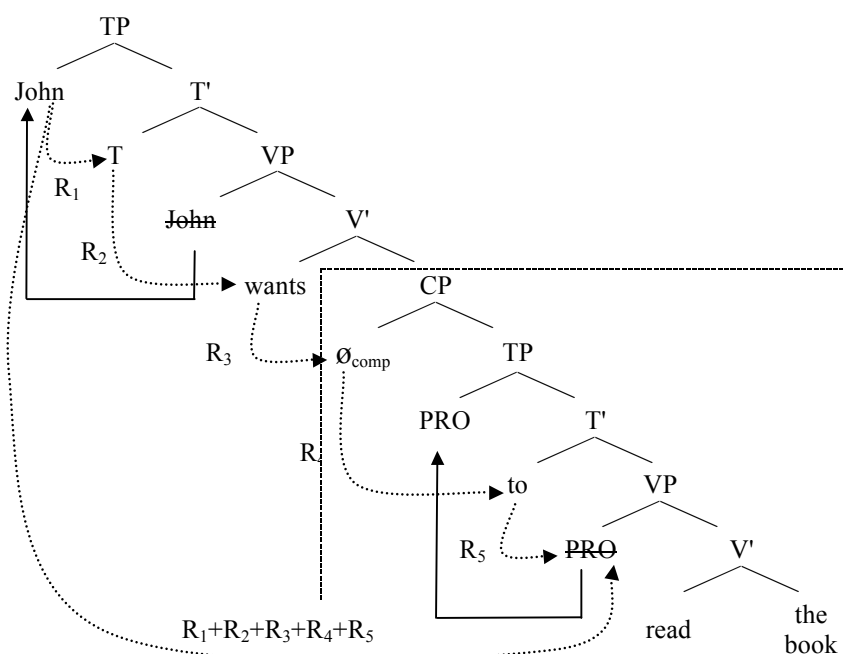


The relations are defined in terms of agree-chains, which is the result of the operation Agree: two or more elements that share a given feature form an agree(ment)-chain. In (86), the uninterpretable φ -features of infinitive T (*to*) make it probe its c-command domain and encounters PRO with unvalued uninterpretable φ -features. A chain is formed in order to share the unvalued φ -features. C enters in the derivation and due to its uninterpretable and unvalued φ -features, it forms a chain with PRO and *to* and they

all share the unvalued φ -features. The matrix v/V enters in the derivation and its unvalued and uninterpretable φ -features make it possible to form a chain now between matrix v , C , to , and PRO . Matrix T forms a chain with matrix v/V due to its uninterpretable T_s feature (recall that T_s enters valued in v/V). *John* and matrix T form a chain because of the uninterpretable φ -features of T (note that *John* merges in the specifier of vP and thence it moves to the specifier of T). Since PRO has unvalued φ -features, the syntactic chain that is formed by the composition of $R_1+R_2+R_3+R_4+R_5$ is translated into A-binding at C-I. OC shows syntactic restrictions like c-command or locality as these are operative in the forming of the syntactic chain.

(86) a. $John_i$ wants $PRO_{i/*j*k}$ to read the book²⁹.

b.



- R_1, R_3, R_4 and R_5 = agree chains $\langle \varphi$ -features \rangle
 R_2 = agree chain $\langle T_s \rangle$
 $R_1 + R_2 + R_3 + R_4 + R_5$ = chain (interpreted as A-binding at C-I)

Since this is the most economical option of encoding the dependency (it requires fewer interpretative steps, see Reuland 2001:473-474), whenever PRO is able to form a syntactic chain, all other options are blocked.

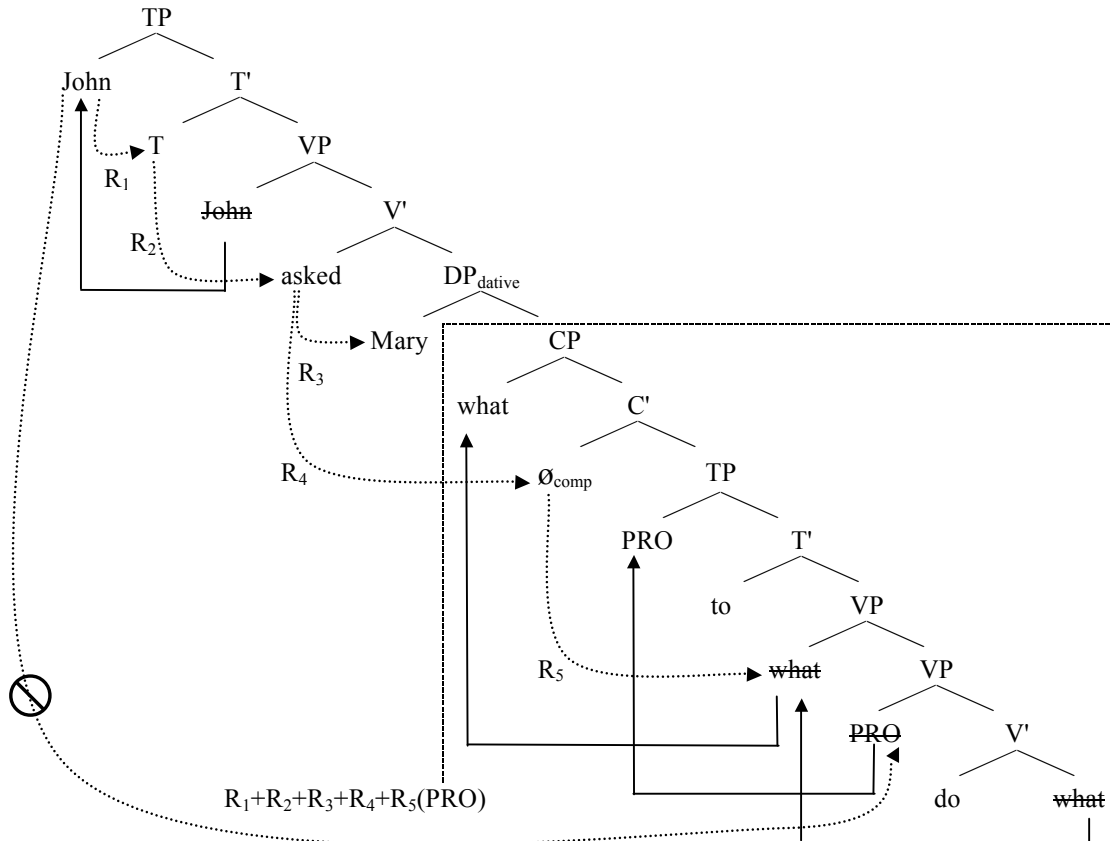
²⁹ I assume in the representation of this sentence that subjects are generated in [Spec, VP], and abstract away from v and To for expository purposes.

3.4.1.2. *Non obligatory control*

When PRO cannot form a syntactic chain with a suitable antecedent, it still may be bound by a precedent DP by means of A-binding directly at C-I (see section 3.3.3). In such a case, PRO behaves like bound pronominals, yielding NOC.

(87) John_i's wife believes [_{CP} that[PRO_i keeping his anger under control] will be good]

(88) a. John_i asked Mary_j [_{CP} what [to PRO_{i/j} do]]



Since PRO in (87) cannot form a chain with any nominal before being interpreted at C-I, it remains with its interpretable φ -features unvalued. At C-I interface, it may still be bound by *John*. In (88), no chain can be formed between *John*, *Mary* and PRO. C° agrees with *what* and not with PRO because of the *wh*-feature in C° (note that *what* and PRO are equidistant once the former has moved to [Spec, vP] so *what* is attracted because it allows the deletion of the *wh*-feature as well as the valuation of the φ -features of C° in one step).

Whenever PRO can be bound neither by means of a chain in syntax nor at C-I, it still may corefer to elements accessible enough in the discourse storage by means of

covaluation at discourse level. This accounts for cases of NOC PRO where the controller does not c-command PRO but the other way around, like in (89).

(89) PRO_{i/*j} reading that book is difficult for me_i

3.4.1.3. Some notes on arbitrary control

When none of the previous options is plausible because there is no antecedent whatsoever, not even in the discourse storage, PRO takes *arb* value as a last resort. This is the most costly option: *arb* is blocked in (90a) because there is a possible controller and this may be done by covaluation at discourse level in the same way as in (89). In (90b) there is, however, no other option than *arb* (assuming that the sentence is completely isolated or the context provides no suitable antecedent). The true value *arb* is only compatible with a totally free pronoun (i.e. not constrained by φ -features) since *arb* may refer to speaker, addressee/s, other/s or any possible combination. If the pronoun had any φ -feature specified³⁰, its interpretive freedom would be constrained. Hence *arb* is only compatible with PRO.

(90) a. Hi John_i. What are you_i doing? PRO_{i/*j} reading that book?

b. PRO_{arb/*j} reading that book is difficult

I will come back to the issue of PRO_{arb} and arbitrary control later on in chapter 7.

To recapitulate so far, PRO is a SE-anaphor different from Dutch *zich* in that it has its φ -features completely unvalued, and has no phonological content. Its interpretation depends on the structural position it occupies. Recall from section 3.1. that economy holds in binding. Whenever the dependency between a potential controller and PRO can be encoded within syntax (by chain formation, yielding OC PRO) this option blocks binding at C-I (which yields most cases of NOC PRO), which in turn blocks coreference by covaluation at discourse storage (which yields the rest of the cases of NOC PRO). Finally, whenever PRO cannot be bound by or cannot corefer

³⁰ Dutch *men* has an *arb* value but I do not consider it a real *arb*. In sentences like (I), *men* usually refers to a group of people where neither the speaker nor the addressee are usually included unless otherwise indicated. This must be due to its person feature specified as 3rd (which can be seen in the agreement morpheme of the verb):

(i)	Men	zeft	dat	jij	een	leugenaar	bent
	men	says _{3rd,sing}	that	you	a	liar	are

to anything at all (there is no suitable antecedent whatsoever in the context), it receives *arb* value.

3.4.2. The distribution of PRO

The hypothesis that I will proceed to defend here is that the distribution of PRO, being a null SE-anaphor, is accounted for like the distribution of any other DP, namely, by the interaction of the Case and Phase Theories (Chomsky 2001, 2006). I will implement this idea by means of two concepts that I will introduce right below: *long vs. short ECM* and *direct vs. indirect ECM*. Besides, I will use the distinction between *structural vs. thematic Case* used by Reinhart & Siloni (2005) in order to describe control infinitive clauses in Spanish.

3.4.2.1. Case and silent PRO

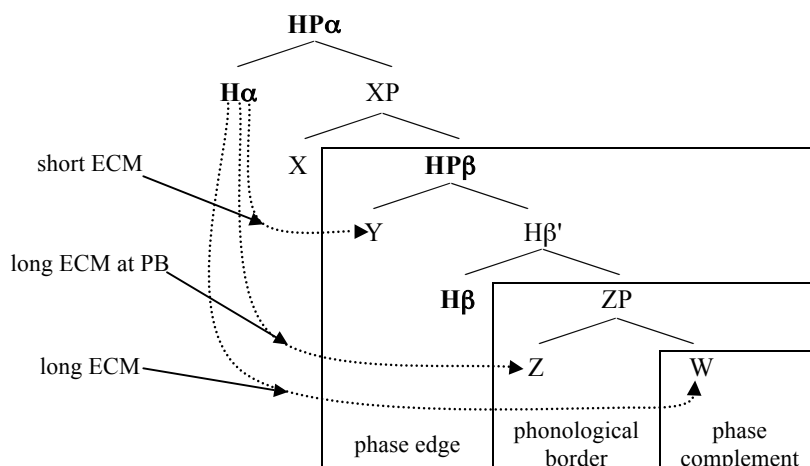
In what follows, I will review two concepts I elaborated elsewhere (Teomiro 2005), which are naturally derived from Chomsky's (2001, 2006) Phase Theory and Pesetsky & Torrego's (2004, 2007) views of structural Case and feature sharing previously discussed. These concepts, Long. vs Short ECM and Direct vs. Indirect ECM, are the way I will implement the idea of interrelation between Case and Phase theories.

Long vs. short ECM

I propose that in a configuration such as (91), formed by two phases, α and β :

- a) Short ECM takes place between the element Y at the edge of the phase β and the element X in the phase α .
- b) Long ECM at the Phonological Border (PB) holds between X in the phase α and Z in the phase β if and only if both Y and H_β have no phonological content (i.e. Z is at the phonological border - defined in (92) - of the phase β).
- c) Long ECM is established between W in the phase β and X in the phase α if and only if Y, H_β and/or Z have phonological content (i.e. Y is not at the phonological border).

(91)



(92) Phonological edge:

(Chomsky 2001:27-28)

An edge element with no phonological material c-commanding it within the category.

(93) The phonological edge of HP is accessible to probe P.

(Chomsky 2001:28)

Direct vs. Indirect ECM

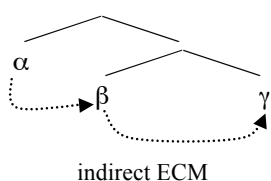
I propose that in a configuration such as (94):

a) Direct ECM takes place between α and γ without β intervening.

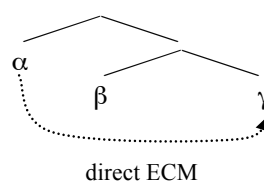
b) Indirect ECM takes place between α and γ when β intervenes.

(94)

a.



b.



For the time being, we can draw some conclusions:

a. In (91), $Z = [\text{PRO}/^*\text{DP}]$ unless Z is at the PB of β

b. In (91), $W = [\text{PRO}/^*\text{DP}]$

c. In (94), α can trigger movement in γ if and only if there is direct ECM³¹

Let us now proceed to analyse the distribution and interpretation of PRO in infinitive clauses both in English and Spanish.

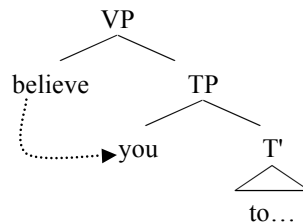
³¹ This is to say that movement is possible if and only if there is a probe-goal relation between α and γ .

3.4.2.2. *English infinitive clauses*

I will now analyze English infinitive clauses and the distribution of overt vs. null subjects with the theoretical tools presented above:

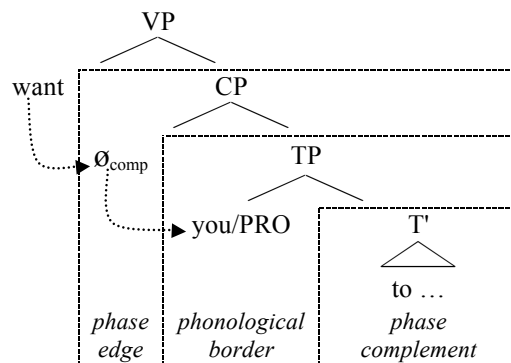
- a) Believe-type ECM: assuming that ECM verbs select TPs (Chomsky 2001, 2006) and that *to* is the morphological realization of infinitival T°, subjects must be licensed by means of direct (since the subject can be extracted by means of A-movement) and short (no phase boundary) ECM.

(95) I believe [TP you to...]



- b) Want-type ECM/Control: assuming that they are CPs (Chomsky 2001) and that *to* is the morphological realization of infinitival T, the subject must be licensed by indirect (it cannot undergo A-movement) and long at the PB (since the subject may be an overt DP besides PRO) ECM.

(96) I want [C° you/PRO to...]

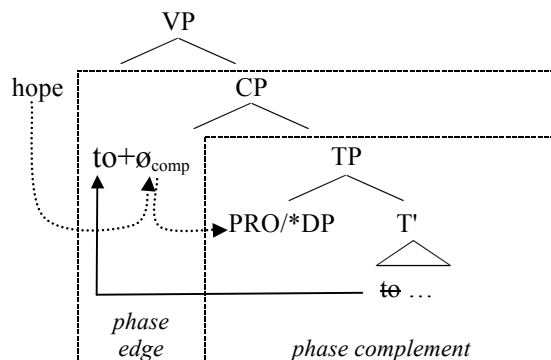


- c) Hope-type Control: assuming that they are CPs, that *to* is the morphological realization of infinitival T and that T-to-C movement has taken place³², only PRO

³² Also Pesetsky (1982) suggests that the relative position of PRO may well be to the right of *to* rather than to its left.

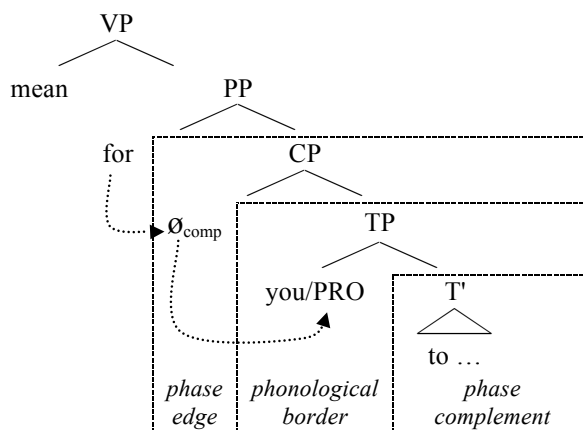
can be licensed at the subject position by means of long and indirect ECM (since C has morphological content, namely the moved tense marker *to*).

(97) I hope [to+C PRO \bar{t} ...]



d) For-to Infinitives: are accounted for in a parallel way as Spanish P+infinitives, discussed in section 4.2.1. *For* is a preposition able to assign structural Case, i.e. it has a [+T] feature. The subject of the infinitive clause, which is a CP, is licensed by indirect and long ECM at PB (infinitival C has no phonological content if we assume that no T-to-C movement has taken place).

(98) I mean [_{pp} for [C° you/PRO to...]



e) ing-Clauses: these are extensively discussed in section 3.5.

3.4.2.3. Spanish infinitive clauses

I will assume that Spanish infinitivals are clauses. However, it is not clear that they constitute strong phases, which can be argued on the basis of the following facts.

On the one hand, some Spanish infinitivals allow clitic raising with restructuring verbs (Rizzi 1982) such as *esperar* ("hope"), *amar* ("love") and *querer* ("want"), as shown in (99) with the verb *querer* ("want").

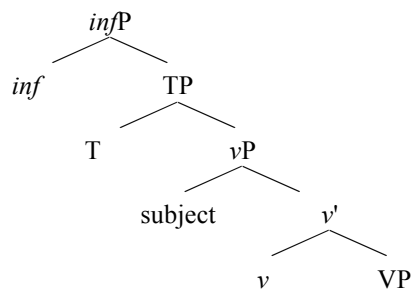
- (99) a. Yo quiero [ir -me]
 I want [go_{inf} -me]
 b. Yo **me** quiero [ir -~~me~~]
 I me want [go_{inf} -~~me~~]

On the other hand, nominals in subject position of the infinitival selected by a preposition can be either anaphors bound by the matrix subject or pronominals as long as they do not refer to the matrix subject (otherwise a Condition B violation emerges) as in (100a). This tells us that a syntactic chain between the matrix subject and the embedded subject can be formed (note that the embedded subject is not at the phonological border so no phase boundary can intervene between the subjects if a chain is to be formed). If the chain between the subjects could not be formed, *ella_i* (bound by the matrix subject) in (100a) would not be blocked.

- (100) a. María_i se descalzó al llegar ella*_{i/j} a casa (based on Fernández Lagunilla 1987:135)
 b. María_i se descalzó al llegar ella*_{i/j}/PRO_i a casa
 (based on Fernández Lagunilla 1987:135)
 María_i se put her shoes off when arrive she*_{i/j}/PRO_i home
 “Maria put her shoes off when she arrived home”

Therefore, I will label the C-like head of Spanish infinitivals *inf*, in a parallel way as I will do with English *ing* in section 3.5.2. Due to the presence of this head, the subject cannot be licensed by means of direct and short (since there is no phase boundary) ECM though it would be possible, in principle, by indirect ECM. The structure of Spanish infinitival clauses is represented in (101).

(101)



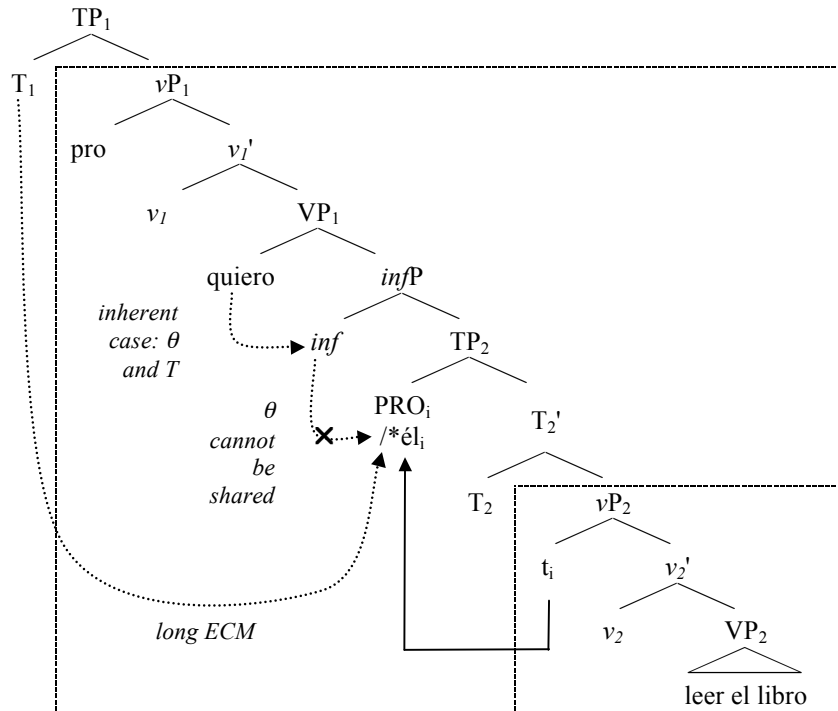
As seen before in (7), repeated below as (102), Spanish infinitival clauses do not allow lexical subjects when they appear as complements to transitive verbs. I will put aside the cases of NcI, which will be discussed in section 3.6.

- (102) a. Yo quiero [**PRO** leer el libro]
 I want [**PRO** read_{infinitive} the book]
"I want to read the book"
- b. *Yo quiero [**María** leer el libro]
 I want [**María** read_{infinitive} the book]
"I want Maria to read the book"

The question is why overt DPs are not licensed in cases such as (102). Following the reasoning of Reinhart & Siloni (2005), I will assume that Spanish transitive verbs do not assign structural accusative but inherent or *thematic* (in Reinhart & Siloni's terms) Case³³. Torrego (1998b) had already proposed that not all transitive verbs in Spanish assign structural Case. If this claim is in the right track, the subject of the infinitival cannot be licensed by indirect ECM because thematic Case is tied to theta assignment. This is to say that the verb does not select the embedded subject but the infinitive clause instead, i.e. the *inf* head. The Case feature of *inf* can be valued but it cannot be shared by the subject because it is not theta marked by the verb. This is represented in (103).

³³ Russian also has control bare infinitives which do not allow overt subjects. However, in the case of Russian, it is more difficult to argue that it lacks structural accusative Case, due to its rich Case system. Another option is to say that verbs, in general, assign structural Case only when they select DPs. When CPs are selected, on the other hand, only inherent Case is available. This question has, nonetheless, numerous ramifications and it is beyond the scope of this thesis.

- (103) Quiero PRO/*él leer el libro.
pro want_{1st,sing} PRO/*him read_{infinitive} the book
 "I want PRO/*him to read the book."



Note that Middle Spanish as well as some literary uses of modern Spanish allow overt subjects in some infinitives³⁴ (Hernanz 1999), as shown in (104) and (105) respectively, like Portuguese inflected infinitives (Raposo 1987). It is worth saying that Middle Spanish verbs are thought to assign structural Case (see Reinhart & Siloni for relevant tests). That structural accusative case is found in certain literary uses of actual Spanish, might be indicating a change in the parameter, i.e. Spanish is changing from a stage in which the transitive verbs assigned structural accusative Case (Middle Spanish), to another stage (Modern Spanish) in which transitive verbs can only assign thematic accusative Case.

- (104) el dulce sonido de tu habla me certifica ser tu mi señora Melibea
 the sweet sound of your voice me_{clitic} certifies be_{inf} you my lady Melibea
 "The sweet sound of your voice certifies to me that you are my lady Melibea"

³⁴ The example in (104) is taken from "La Celestina" (S. XVI) by Fernando de Rojas, cited by Hernanz (1999). Hernanz also notes that certain modern uses of the language allow this kind of constructions, as shown in (105) from "Una comedia ligera" by E. Mendoza (1996).

- (105) ... Sancha García advirtió estar la puerta del salón abierta ...
 ... Sancha García realized be_{inf} the door of the living room opened ...
 "...Sancha García realized that the door of the living room was opened..."

So how is the subject licensed? When the matrix v enters in the derivation, which is a strong phase head, the phase complement must be sent to the S-M system. The Case of *inf* is not a problem since it is already valued. However, the Case of the embedded subject has not been valued yet. Therefore, if the subject is an overt DP the derivation crashes at PF. Conversely, if the subject is PRO, since it has no phonological values whatsoever, its lack of Case is tolerated at PF and it may be licensed by means of long ECM by the matrix T° . Besides, it forms a chain with the matrix subject, which yields OC.

To sum up, the phonological status of PRO allows it to be licensed in environments where overt DPs are not because they have to be spelled out. Since Spanish transitive verbs do not assign structural Case, indirect ECM with matrix v is not an available mechanism to license the subject of infinitive clauses that are complements to transitive verbs. The next available head that can assign structural Case is matrix (finite) T° . However, by the time it enters into the derivation, the material in the embedded subject position has already been sent to the S-M system and the derivation crashes at PF due to the lack of morphological specification for the structural Case of the subject. This is not a problem in the case of PRO as it has no phonological specification whatsoever and its lack of valued structural Case can be tolerated at PF. PRO is eventually licensed by Long ECM with matrix T° .

3.5. English *ing*-clauses

In this section, two analyses of English *ing*-clauses³⁵ will be compared: one by Pires (2001), which draws upon Hornstein's MTC; and the other one will be developed here

³⁵ Reuland (1983) notes points out that the morpheme *-ing* can be found in a variety of constructions like the progressive tense of the verb ("*What are you doing*"), participles that modify a verb ("*I got into the office thinking on that*"), serving as nominalizer ("*The killing* upset the general") or in nominals called POSS(essive)-*ing* by Reuland (1983) ("*Mary recalled his borrowing of the book*"). There is another construction in which the verb in its present participle form takes a nominative subject: the nominative absolute construction ("*Mary didn't want to talk to John, he being confirmed to be the murderer*"). This section focuses on none of these constructions but only on the clauses that Reuland called NP-*ing*, which occupy argument positions ("*I hate Mary singing that song*"; "*Mary singing that song was horrible*"),

and draws upon Pesetsky & Torrego's (2004, 2007) view of feature valuation and of Case as uninterpretable Tense. Since I have not found empirical counter-evidence to either analysis, I will compare them on grounds of elegance and simplicity; i.e. I will assess which of the two approaches can account for the distribution of *ing*-clauses with the fewer theoretical stipulations.

3.5.1. Pires's (2001) analysis

Following the theory of control proposed by Hornstein (1999), Pires proposes that OC PRO is the result of A-movement. He extends this analysis to the case of *ing*-clauses³⁶. They constitute an interesting case due to their ambiguous syntactic properties. Namely, on the one hand they behave as full clauses (they allow long *Wh*-extraction and they have a canonical subject position) but on the other hand they need to appear in Case positions, like DPs. This kind of clauses has several syntactic properties that make them different from *believe*-type ECM (Reuland 1983):

- i. The subject may alternate between a lexical DP and PRO
- ii. The subject appears in morphological accusative case when the clause is complement to a verb or preposition and when it is in subject position.
- iii. Anaphors can occupy the subject position of the *ing*-clause as long as the matrix subject binds them (Canonical Condition A), whereas the matrix subject cannot bind pronominals (Canonical Condition B).
- iv. Long *Wh*-movement of both object and subject is allowed
- v. Short *Wh*-movement is ruled out
- vi. A-movement of the subject into the matrix clause is disallowed (unlike "standard" ECM cases)³⁷

where NP refers to both a lexical NP and PRO ("*I heard PRO singing that song yesterday*"). This latter kind of clauses are what I will be calling *ing*-clauses throughout the present thesis and Pires (2001) labels Clausal Gerunds (CG) in his work.

³⁶ Although Pires used the term *Clausal Gerund (CG)* in his original work, I will be using the term *ing-clause* throughout this thesis for the sake of simplicity.

³⁷ However, perception verbs do allow A-movement from *ing*-clauses. It has been proposed that these verbs may select small clauses instead of "pure" *ing*-clauses as their complements (Reuland 1983, footnote 7). Pires (2001) argues that *ing*-clauses selected by perception verbs (among some other verbs) differ from "standard" cases of *ing*-clauses in that they lack the T layer, i.e. they are *vP* (possibly with a

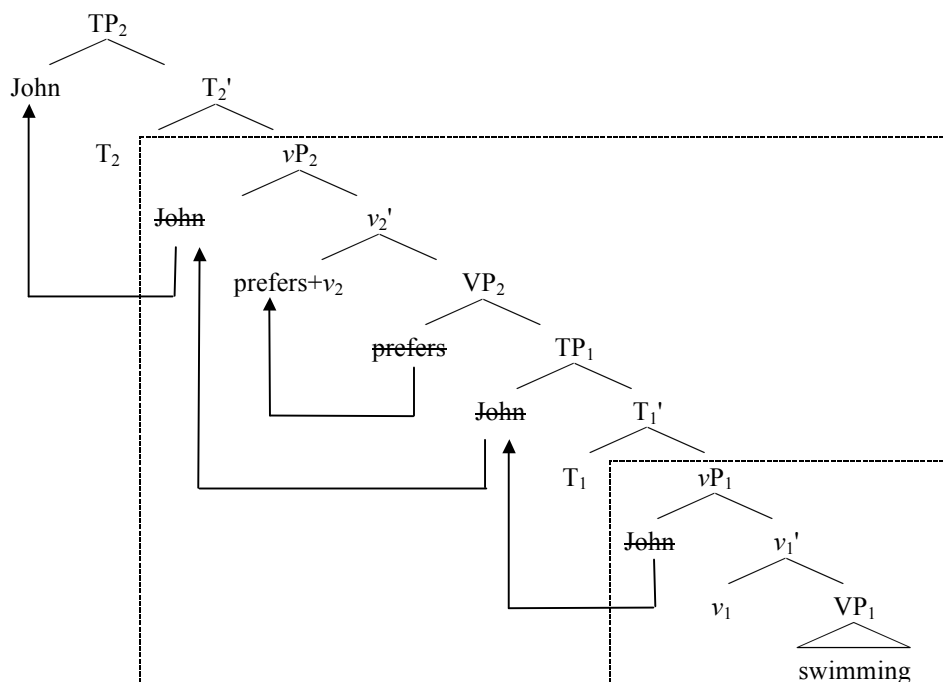
Based on the facts that *ing*-clauses never show an overt complementizer and that they disallow short *Wh*-movement, Pires argues that this kind of clauses are TPs rather than CPs. Besides, he makes the following assumptions in order to account for the properties and distribution of the *ing*-clauses:

- a. the head of the *ing*-clause, namely T, bears an uninterpretable Case feature (in the sense of Chomsky 2001). This is how Pires implements the idea of the head of *ing*-clauses as having a nominal nature.
- b. T has an EPP feature.
- c. T enters in the derivation φ -defective.
- d. When DP merges in [Spec, TP], it transfers its φ -features to T
- e. T can check the Case feature of DP if and only if its own Case feature has been checked

Let us briefly see how this analysis derives the contrast between (106) and (107) (examples and derivations from Pires 2001):

(106) John_i prefers PRO_i swimming

(Pires 2001:31)

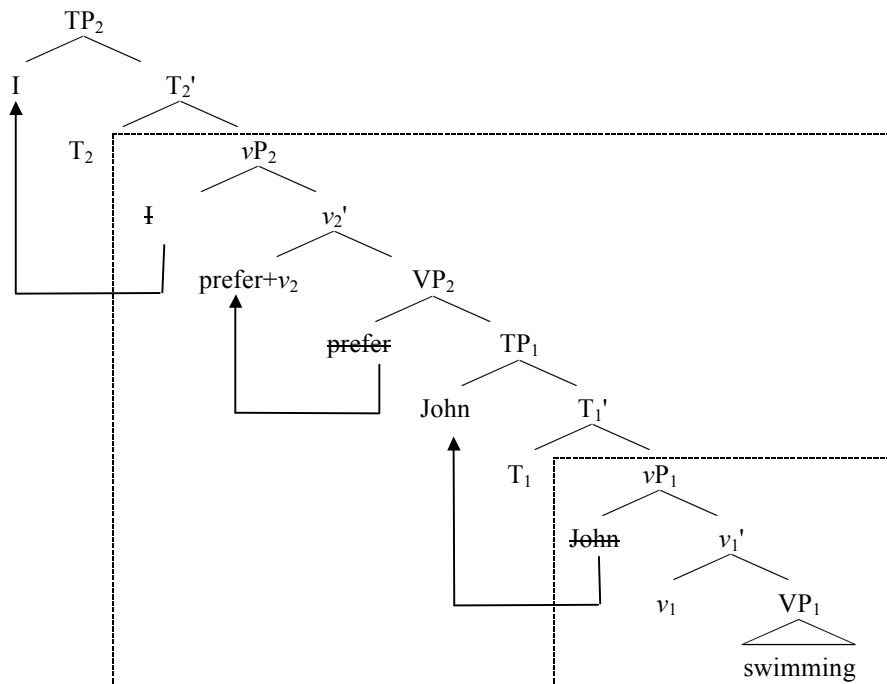


higher aspectual head). Along with perception verbs, *remember* also allows A-movement of the embedded subject:

- (i) He was *seen* doing physical exercises every day
- (i) He was *heard* telling the truth
- (ii) He was *remembered* telling the truth

(107) I prefer John swimming

(Pires 2001:29)



The central point of the analysis is that T_1 cannot check the case of DP in its specifier unless its own Case feature has been checked. The element that will check this Case feature is either matrix v or matrix T in the case of *ing*-clauses in subject position.

In (106), the EPP feature of T_1 makes *John* move to [Spec, TP] and there, it can transmit its φ -features to T_1 . Recall that T_1 was φ -defective when it entered in the derivation and therefore it cannot check the Case of *John*. Now, even with its transmitted complete φ -set, it cannot check the Case of *John* because it has its own Case feature unchecked. At this point, both T_1 and *John* are active for further computations because they both have unchecked features. When matrix v enters, it attracts *John* to its specifier and assigns it a second theta-role (recall that theta-roles are features under the MTC). However it does not assign it Case. Matrix v further agrees with T_1 and checks the unvalued φ -features of matrix v and the Case feature of T_1 . *John* further moves to matrix [Spec, TP] to check the EPP feature of T_2 and check its own Case feature.

The difference between (106) and the derivation of (107) is the timing of the attraction by matrix v : the derivation proceeds as in the previous one until matrix v enters. It first agrees with T_1 and checks its Case feature (recall that matrix v is φ -complete). Now, T_1 agrees with DP, which moves to the specifier to check EPP and

there, it transmits its φ -features to T_1 which in turn can check the Case of the DP since its own Case feature has already been checked.

This analysis has two shortcomings:

- a) **The clausal status of ing-clauses:** that ing-clauses are TPs instead of CPs is not clear at all (see arguments in Reuland 1983:107-111).
- b) **The relation between the Case feature of T and its ability of checking the Case feature of the subject:** it is not clear how in (107) T_1 can check the Case feature of *John*. The DP has transmitted to T_1 its φ -features and now T_1 is no longer φ -defective. However, T_1 needs to agree again with DP so that its inherited (from DP) complete φ -set can check the Case feature of DP. This seems to be a new mechanism of Case checking under the framework of Chomsky (2001). It is not clear whether there is φ -features transmission. On the other hand, Chomsky's *Inactivity Condition*³⁸ argues against there being two sequential agree operations between the same two elements. Finally, it is unclear why the unchecked Case feature of T_1 prevents it to check the Case of DP if its φ -features are now (after φ -transmission from DP) not defective any more. Under the system of Chomsky (2001), the checking of the Case feature of a DP rests only upon φ -checking. It is not clear whether the φ -transmission that Pires advocates can be considered φ -checking but even if that were the case, Case should be checked on DP, independently of the Case feature that the head T_1 has.

3.5.2. The alternative analysis

Reuland (1983) gives evidence supporting the idea that *ing*-clauses behave like clauses: unlike nominals, they cannot move to the left periphery of the sentence as in topicalization. Besides, *Wh*-extraction is allowed, which would not be expected if *ing*-clauses were DPs. The obligatory presence of a subject in the clause (either an overt DP or PRO) reflects the EPP, which holds for sentences (clauses) but not for nominals. However, these clauses retain some nominal characteristics, e.g. the tense of this kind of clauses is not "strong" enough so as to license a subject internally, i.e. without the help of an external element. Furthermore, the nominal character of *ing*-clauses can be seen in

³⁸ See Nevins (2004:9).

sentences such as (108), where the subject bears genitive case (typical of nominals) and not accusative/nominative (typical of clauses).

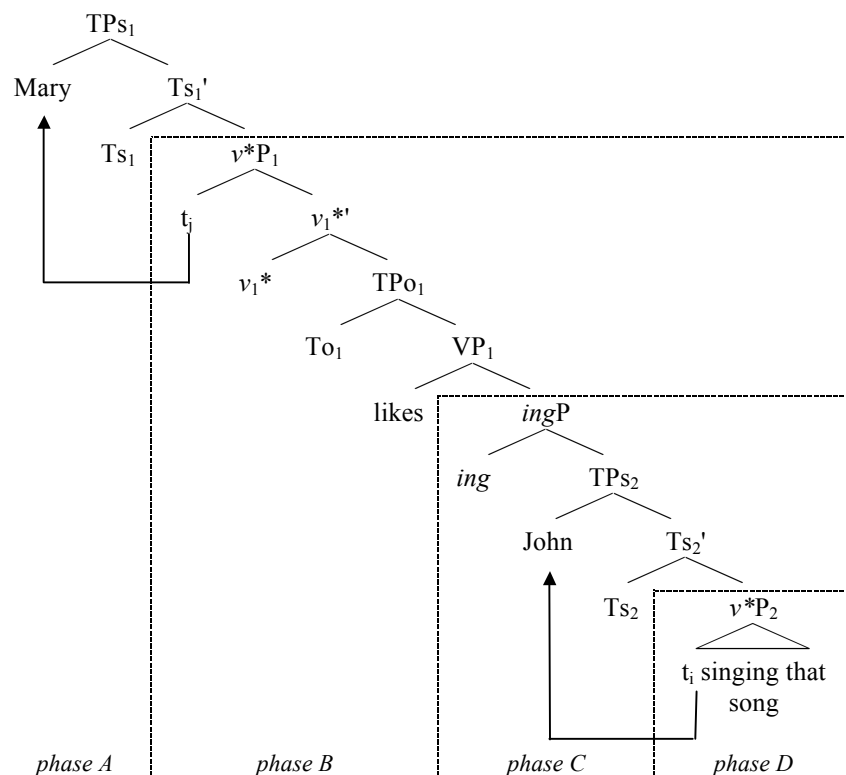
(108) Which movie would you disapprove of [_{ingP} my seeing]? (Reuland 1983:108)

I will consider *ing*-clauses as full clauses, i.e. as having a C-layer, projected by a functional head dominating the tense node of such clauses. I will label this head *ing* in order to put aside the question of its precise nature. This functional head *-ing-*, higher than the subject in the syntactic tree, is the element that mediates between the embedded subject and the element that licenses it (e.g. a matrix *v* or *T*^o). Nonetheless, assuming the presence of this functional head, the question raises as to whether this head defines a strong phase or not. For now I will put aside this point until I discuss the phonological status of the *ing* head.

I will show the derivation of both *ing*-clauses and *believe*-type of ECM in order to explain what the differences of the two follow from.

In *ing*-clauses there is no probe-goal relation between the matrix verb and the embedded subject. Note that C (in this case, our functional head *ing*) has φ - and tense features (Chomsky 2006), though they are both uninterpretable on C. The structure of a sentence with an embedded *ing*-clause is represented in (109).

(109) Mary likes [_{ingP} John singing that song].



In (109), the tense of the *ing*P is introduced by v^*_2 , but it is unvalued, unlike in tensed clauses. When T_{S_2} merges with v^*P_2 , its unvalued interpretable T feature makes it peruse its c-commanding domain so as to value this feature. It encounters the DP "John", with which it agrees and forms an agree-chain sharing the (still unvalued) T feature. T_{S_2} keeps searching and finds v_2 , which has another instance of T. They both agree and now there is an unvalued agree-chain $\langle T_{S_2}, \text{John}, v_2 \rangle$. There is an instance of valued T below in the tree, namely in the head T_0 within v^*P_2 . However, it is within a closed phase and therefore, it is unable to value the T of the so far formed agree-chain.

The derivation proceeds and *ing* merges with TP_{S_2} . Its uninterpretable φ and T features make it peruse its c-command domain and it probes the DP "John", which has moved to $[\text{Spec}, TP_{S_2}]$ due to the EPP feature of T_{S_2} . Agree applies between them and now there are two agree-chains: one sharing unvalued T $\langle \text{ing}, \text{John}, T_{S_2}, v_2 \rangle$ and the other one sharing valued φ $\langle \text{ing}, \text{John} \rangle$.

The matrix verb V_1 enters, which has a valued T feature (that will be interpreted in T_{01}). Its unvalued φ -features force it to explore its c-commanding domain and it encounters *ing*, which has uninterpretable but (by now) valued φ -features³⁹. Agree applies and the agree-chain of valued φ -features is composed by $\langle V_1, \text{ing}, \text{John} \rangle$. Due to the relation established between it and *ing*, T on V_1 is able to value T on *ing* and thus in the whole agree-chain that shares T, which is now composed by $\langle V_1, \text{ing}, \text{John}, T_{S_2}, v_2 \rangle$. The connection between the matrix verb and *ing* due to their φ -features, is able to value T on the *ing*P⁴⁰. This means that structural case (uninterpretable T on DP "John") is assigned and, consequently, DP "John" is licensed.

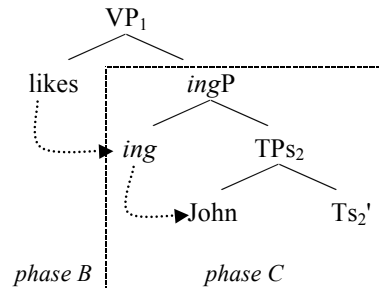
There is an agree-chain $\langle V_1, \text{ing}, \text{John}, T_{S_2}, v \rangle$ that shares T. However, there is no probe-goal relation between $\{V_1, \text{John}\}$ (there is one probe-goal relation $\{V_1, \text{ing}\}$ and another between $\{\text{ing}, \text{John}\}$, as represented in (110), which allows V_1 and DP "John" to share the feature T). In other words, "John" is licensed by indirect ECM in (110). No movement of the DP "John" can be triggered by the head that values its case, as matrix

³⁹ Following Chomsky (2001:13), uninterpretable features do not delete until the next higher phase is closed.

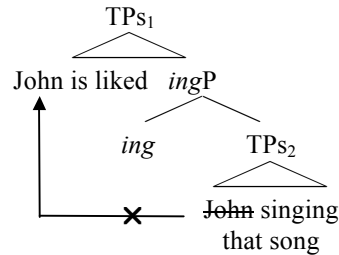
⁴⁰ By means of the Match Condition (Pesetsky & Torrego 2004). Pesetsky & Torrego (2007) manage to account for the licensing of nominative subjects dispensing with the Match Condition but say nothing about accusative objects

T° in (111) where the matrix passive verb is unable to assign structural Case, since a probe-goal relation between them is crucial for movement to take place (Chomsky 2001).

(110)



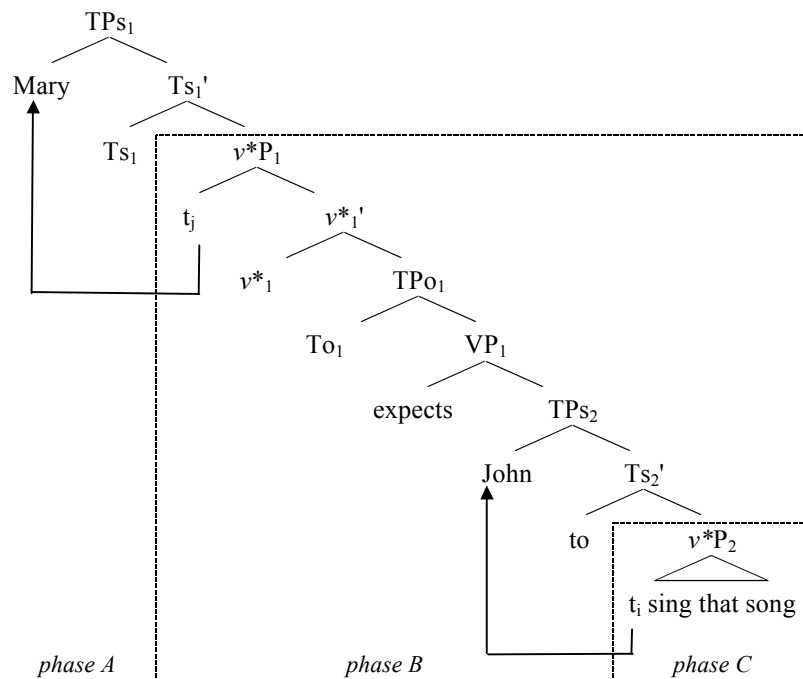
(111) *John is liked singing that song.



In ECM clauses there is probe-goal relation between the matrix verb and the embedded subject. If we look at the structure of an ECM construction (this being considered a bare TP⁴¹ (Chomsky 2001, 2006)), it can be seen that the agreement between the matrix verb and the subject is direct, i.e. without any intervening element.

⁴¹ We can also think of ECM configurations as CPs. It could be said that the subject moves to [Spec,CP], whence it can be assigned case. The empirical results are identical since the first element that the matrix verb "encounters" in order to get valued its ϕ -features, is the DP and not C; so that a probe-goal relation between the matrix verb and the DP takes place, unlike in *ing*-clauses.

(112) Mary expects [TP John to sing that song]

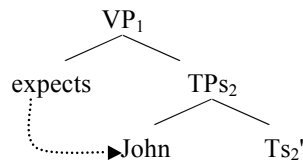


The unvalued interpretable T feature forces T_{S_2} to peruse its c-commanding domain and it encounters the DP "John". Agree applies and an agree-chain sharing the unvalued T is formed between $\langle T_{S_2}, \text{John} \rangle$. T_{S_2} keeps searching and finds v^*_2 , but its T is unvalued. Nonetheless, it forms part of the unvalued agree-chain $\langle T_{S_2}, \text{John}, v^*_2 \rangle$.

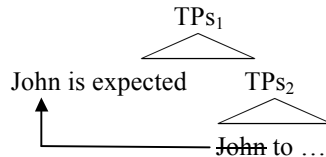
When the matrix verb enters into the derivation, its unvalued φ -features make it look at its c-commanding domain in search for an element that has valued φ -features. It encounters the DP "John", which has moved to [Spec, TP_{S_2}] by now due to the EPP feature of T_{S_2} . Agree applies and an agree-chain is formed between $\langle \text{expect}, \text{John} \rangle$ to share the φ -features of DP "John". Therefore, the T of the matrix verb becomes "visible" to the DP "John" and thus T gets valued in the whole embedded TP.

Unlike in *ing*-clauses, there is a probe-goal relation between the matrix verb and the embedded subject, as shown in (113). Therefore movement is not prevented when it is triggered by the functional head that values the Case of the embedded subject, as in (114), where the matrix T assigns structural Case (since the passive matrix verb cannot assign structural Case) to the embedded subject and this moves into the matrix cause.

(113)

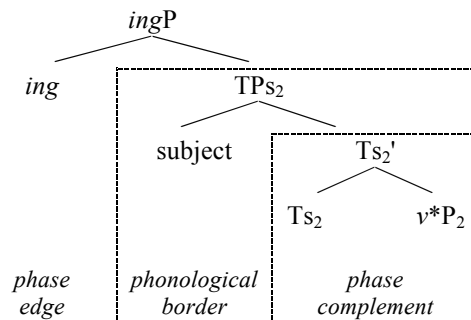


(114)



Silent *ing*: what would happen if *ing* were a phase head? Unless the subject moved to [Spec, CP], it would be within a closed phase by the time the matrix verb merged with the *ing*P. The subject would not have any value for case and the derivation would crash at PF. However, this is not a problem given that - even if the subject does not move to [Spec,CP] - the subject is in the phonological border (defined in (92)), as can be seen in (115).

(115) Phonological border of *ing*P:



Given that an element within the phonological border is a goal accessible to probes outside the phase, this material cannot be sent to the S-M system along with the material included in the closed phase. Otherwise, it would not be accessible for elements external to the phase. In other words, elements within the phonological border are not sent to the S-M system until the next higher phase.

Indirect questions in *ing*-clauses can give us a piece of evidence supporting that we might be in the right track: short *Wh*-movement in *ing*-clauses is impossible:

(116) *Rudy didn't remember [_{ingP} what_i [Mary doing t_i]] (based on Reuland 1983:112)

This makes sense by saying that the subject has to be in the phonological border in order to receive its case: if the *Wh*-phrase is in [Spec, *ingP*], then all lower positions are no longer in the phonological border. The subject must be sent to the S-M interface without any value for case and the derivation crashes at PF. However, long *Wh*-movement is not problematic because a trace (i.e. a copy without phonological component) is left in [Spec, *ingP*], which does not affect the size of the phonological border.

As for the PRO/DP alternation in subject position of *ingP*, it has already been shown how lexical DPs are allowed to surface in subject position within *ing*-clauses by means of indirect Case assignment. As neither *ing* nor the infinitive have any valued φ -feature, PRO is allowed in subject position as well. If there were valued φ -features, PRO would be constrained by them. The valued φ -feature would render PRO into *pro* or a personal pronoun (see section 3.7.1). The DP/PRO alternation in *ing*-clauses posed a problem for earlier accounts of the distribution of PRO because they considered PRO incompatible with nominative Case. As shown in this analysis, PRO may be Case assigned in the same way lexical DPs are. The only condition for its presence is the lack of valued φ -features that can constrain it. OC shows up when PRO forms a chain with the matrix subject.

3.6. Spanish infinitives: NcI vs. PRO

Ortega-Santos (2002) makes a unified analysis of the environments where infinitivals can appear with lexical subjects. He notes that these infinitivals are not L-marked by the main verb⁴². He proposes that the parameterized version of the EPP⁴³ of Harley (2000) and an agreement absorption mechanism by the main verb, account for the distribution of PRO and DPs in infinitivals: infinitives have abstract agreement in Spanish but a matrix verb absorbs this agreement. Control verbs absorb the agreement of the infinitival complement so that they cannot check the EPP by means of V-Raising (see Alexiadou & Anagnostopoulou 1998) and a null subject (i.e. PRO) is needed to satisfy the [+null] EPP feature. When the verb is not complement of a matrix verb, the abstract

⁴² As long as we assume, as Alexiadou & Anagnostopoulou (1998) do, that preverbal subjects are in A'-positions in NSL such as Spanish. This is not clear however (see Zubizarreta 1994, Cardinaletti 2004, Chomsky 2006).

⁴³ Finite T bear a [+overt] EPP feature whereas non-finite T bears a [+null] EPP feature.

agreement is not absorbed and it checks the EPP feature of the infinitive, so there are no EPP effects and both overt and null subjects are allowed.

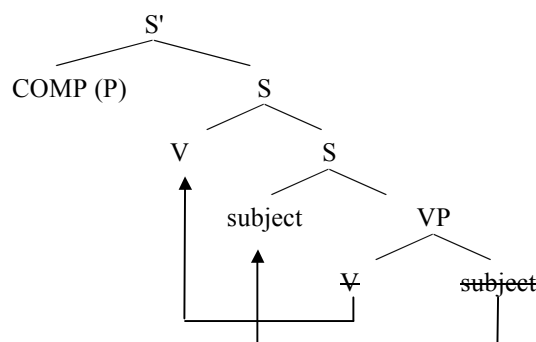
I will not follow this analysis for several reasons. First, it is not clear why the agreement absorption takes place, i.e. why such a mechanism should exist in the grammar. Second, the parameterised EPP seems to be a mere description of the distribution of PRO. Why does finite T need to check its EPP by means of overt subjects whereas non-finite T needs to do so by means of null subjects? What is exactly the relation between finiteness and phonological realization of nominals? In the analysis I propose below, I show that no relation whatsoever holds between finiteness and phonological specification of the nominals but it is the presence/absence of overt valued ϕ -features (overt agreement) that decides whether PRO is allowed or not. From my analysis it follows, moreover, why this is so.

3.6.1. P+infinitive

To my knowledge, Fernández Lagunilla (1987) was the first analysis of this kind of constructions in Spanish under the framework of Government & Binding, where the subject not only needed to be Case-assigned but also governed. She proposed that P+infinitives are clauses with adverbial properties, where the verb has adjoined to the node S (TP in our terms) in order to govern (assuming, along with Torrego 1984, that the verb can govern from an adjunct position) the subject (the verb is governed by COMP, which is lexically filled by P(reposition), because INFL is missing). However, since the subject is not in object position, it cannot be assigned accusative by the verb and neither can it be Case marked by the preposition in COMP, since it was supposed that prepositions assign inherent Case and the subject is not selected by the preposition. It is by means of a default Case strategy, based on work by Belletti (1987), that the subject receives nominative case: *assign nominative to an element directly governed by S*, as represented in (117):

(117)

(based on Fernández Lagunilla 1987)



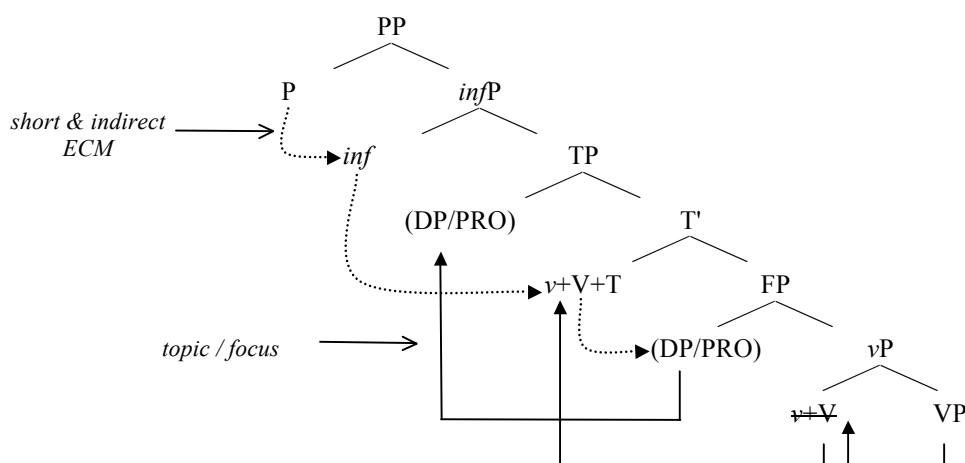
In the analysis I will develop here, I will assume without further discussion that prepositions are Case assigners. In particular, I assume they assign structural Case in Spanish (unlike verbs, which assign inherent Case in this language). This means that P has a [+T] feature that will license the embedded subject (either overt or null, i.e. PRO), as discussed in section 3.6.1.

The embedded subject usually remains in middle field (MF) position (Cardinaletti 2004, Zubizarreta 1994), a fact previously analyzed as VP-preposing (Torrego 1984, Fernández Lagunilla 1987). However, the subject may sometimes raise to the topic position [Spec,TP] as in (118a), though this option is by no means generally available, as can be seen in (118b).

- (118) a. Después de (él) decirlo (el), sé que es verdad.
 After of (he) say_{inf-it-clitic} (he), know_{1st,sing} that is_{3rd,sing} true
“I know it is true after him having told me”
- b. Salí de la habitación al (*Juan) entrar (Juan).
 Got_{1st,sing} out of the room when (*Juan) got_{inf} in (Juan)
“I got out of the room when Juan got in”

The subject is licensed by means of short (no phase boundary is crossed) and indirect ECM, which is possible since P assigns structural Case⁴⁴.

(119)



⁴⁴ See also Rigau (1995).

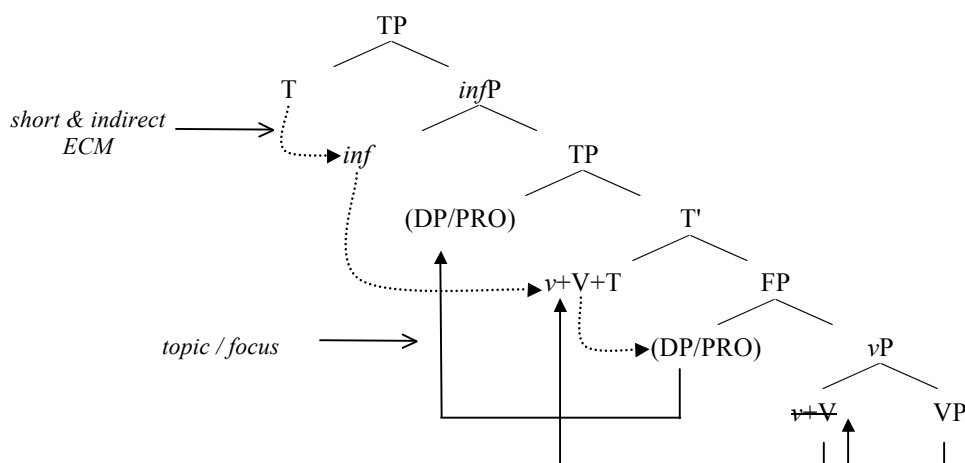
The derivation takes place as follows: once we have the embedded νP , this is merged with the embedded infinitival T° , whose unvalued φ -features make it search for a possible goal. It then encounters the embedded subject.

- a) If it is a DP, it values the φ -features of the infinitival T, which, in its turn, cannot value the T feature of the DP since it is unvalued for T. However, an agree-chain is formed sharing the unvalued feature T. The derivation proceeds and *inf* merges with TP. Its unvalued φ -features make it probe its c-command domain and it encounters T with valued φ -features. An agree chain is formed to share the valued φ -features and another one to share the unvalued T feature. At this moment P enters in the derivation with its unvalued φ -features and its valued T feature. It probes *inf* with valued φ -features and enters in the agree chain so that its φ -features are valued. By means of the Matching Condition, *inf* can see the valued T feature of P and so, T gets valued in the agree-chain, which means that its uninterpretable instances in *inf*, and T are deleted.
- b) If the embedded subject is PRO, T and PRO share the unvalued interpretable φ -features of PRO. Although they are unvalued, the uninterpretable instance can now be deleted in T. The derivation proceeds in the same way as in the case of overt DP. PRO is allowed since the infinitive does not have valued φ -features when PRO enters into the derivation. Its Case is assigned by P and OC PRO arises by forming a chain with a c-commanding DP in the matrix clause (*pro* can also be a potential controller), by means of which the φ -features of the agree-chain formed by <controller, P, *inf*, T and PRO> are valued.

3.6.2. Infinitive in subject position

The subject is licensed by means of indirect short ECM with the matrix T° , just in the same way as in the case of P+infinitives previously described.

(120)



These infinitivals can also be introduced by the determiner *el*, as in (121a) and (121b)⁴⁵.

(121) a. Ir yo mañana a la universidad será difícil.

(Fernández Lagunilla 1987:125)

Go_{inf} I tomorrow to the university will be difficult

b. (El) ir yo a la universidad será difícil.

The go_{inf} I to the university will be difficult

c. *Éste ir yo a la universidad será difícil.

This go_{inf} I to the university will be difficult

d. El despertar de Melkor destruyó la Tierra Media.

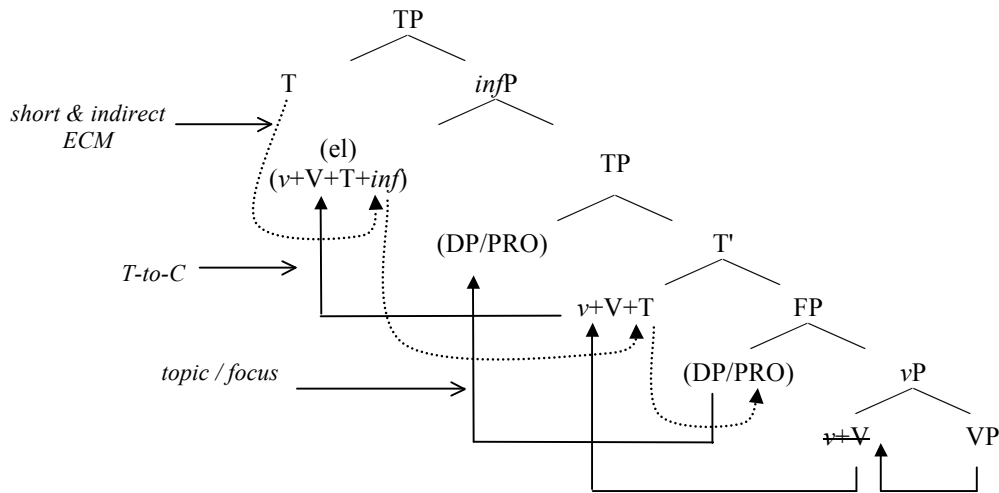
The awaken of Melkor destroyed the Middle Earth

El is the definite article in Spanish. It also introduces nominal infinitivals as in (121d), whose subject must be either in genitive or preceded by the preposition *de*. However, in the infinitivals studied in this section (with nominative subjects), *el* does not seem to be a standard determiner since its position cannot be occupied by a demonstrative, as (121c) shows. I will assume that *el* is the phonological realization of *inf* when T-to-C has taken place⁴⁶, as represented in (122).

⁴⁵ See Serrano Pardo (2008, forthcoming) for a more detailed analysis of clauses introduced by *el* in Spanish.

⁴⁶ It is widely assumed that left adjunction of T° to C° usually gives a complex head with the form T°+C°. However, I will assume that the complex head T°+*inf* is spelled out as C°+T° due to morphological properties of Spanish *inf*. Hence the word order *el*+infinitive.

(122)



This accounts for (123), where the subject can precede the verb only when *el* is not present. When *el* introduces the infinitival, it must be adjacent to the verb and the subject cannot intervene.

(123) a. (?Yo) **ir** mañana a la facultad va a ser difícil.

I_{nom} **go**_{inf} tomorrow to the faculty will be difficult

“It will be difficult for me to go to the faculty tomorrow”

b. **Ir** (yo) mañana a la facultad va a ser difícil.

c. (E1) **ir** (yo) mañana a la facultad va a ser difícil.

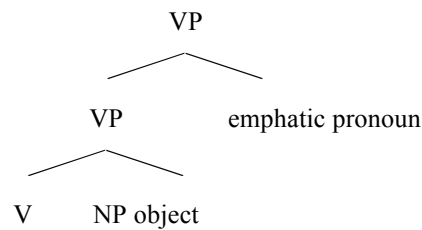
d. (E1) (*yo) **ir** mañana a la facultad va a ser difícil.

Piera (1987) analyzes infinitivals in subject position as having null subjects. The overt pronoun (he assumes that the cases of overt nominals different from personal pronouns are exceptional in this kind of infinitivals) is not a real subject but a nominative⁴⁷ emphatic pronoun adjoined to the VP, yielding a structure such as (124):

⁴⁷ Piera argues that the emphatic pronoun bears nominative Case following a proposal by Burzio (1986): *an anaphor in an A'-position bears nominative Case when it is coindexed with a subject that c-commands it.*

(124)

(Piera 1987:163)



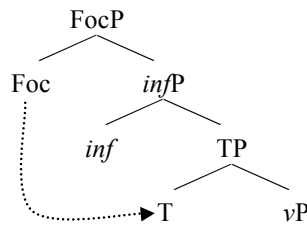
PRO in embedded clauses occupying matrix subject position has traditionally been analyzed as NOC PRO. If PRO is a SE-anaphor with unvalued φ -features, its interpretation would be constrained by the valued φ -features of the emphatic pronoun (assuming that it is adjoined to VP and PRO is in MF position or higher) so that PRO would become pro. But such a complex mechanism can be avoided in my analysis, since the insertion of an overt pronoun in subject position is allowed and short indirect ECM by matrix T° can take place as previously shown.

3.6.3. Exclamative and interrogative infinitives

These infinitive clauses can surface with overt subjects. However, personal pronouns are far more frequent than referential DPs, which has made authors like Fernández Lagunilla (1987) think that this not a matter of Case but of focus.

I will basically adopt the same position. Following Torrego (1998) and Zubizarreta (1994), I will assume that structural Case is contrastive in Spanish⁴⁸. Sevdali (2005) proposes that Ancient Greek infinitives have a Case feature that is normally inactive. Only an active focus feature on C can activate the Case feature of infinitival T. This Case has an emphatic function as it emphasizes the contrast between the individual the subject refers to and another entity in the context. I propose that the same happens in Spanish interrogative and exclamative infinitivals: the Case that the subject bears is a contrastive Case with emphatic function activated in T by a focus feature in the left periphery of the clause.

⁴⁸ According to Sevdali (2005), the main idea behind this proposal (the notion of contrastive case) is that *"while Case is available on T, it is only activated in AG infinitival clauses when Focus is at play in the left periphery [...]. In all other cases [...] C does not project at all and T's case feature is not activate / is suppressed."*

(125) Left periphery of an exclamative/interrogative infinitival in Spanish:

Case is valued by means of Foc activating the [+T] feature of infinitival T. An overt DP is licensed as well as PRO, which also needs Case to be licensed under the analysis developed in this thesis. The lack of valued φ -features, allows the presence of PRO, whose interpretable (yet unvalued) φ -features delete the uninterpretable instances of φ -features in *inf* and T so that the derivation can converge.

3.7. Notes on some other issues

In this section I will review a number of cases to which the proposed analysis could potentially be extended. First, the empty category *pro* is analyzed as an instantiation of null SE-PRO with overt φ -features in T^0 . Subsequently, constructions with control into finite clauses in Hebrew, Finnish and Brazilian Portuguese are addressed, as well as inflected infinitives in European Portuguese. Afterwards, OC into adjuncts clauses is discussed.

3.7.1. Little *pro*

Let us assume that apart from PRO, there is another null element available in UG: PRO*. Let us further assume that this element has a full set of φ -features, which renders it a pronominal rather than a SE-anaphor. Then, little *pro* reduces to a specific instantiation of PRO* when it enters into an agree relation with a T^0 with valued φ -features and valued T (e.g. finite T). This idea essentially follows the basic intuition put forward by Rizzi (1982, 1986): *pro* enters without any value in the derivation and by means of the Agr features of INF, it gets licensed and identifiable.

Recall from section 3.4 that PRO is a DP with unvalued & interpretable φ -features and an unvalued & uninterpretable T feature (structural Case). PRO* is similar to PRO except for the fact that it has a full set of φ -features. PRO is merged within vP , like any other subject, and T, due to its unvalued uninterpretable φ -features or to its unvalued T feature (in the case of infinitives) probes for a goal and encounters PRO (or PRO*).

Infinitival T has no φ -features, or these are not valued (infinitives in languages like English and Spanish have no overt agreement features). If infinitival T encounters PRO, two agree-chains form between them both: one to share the unvalued φ -features (if we assume that infinitives have abstract agreement) and the other one to share the T feature that is also unvalued. The EPP of T forces PRO to move to [Spec, TP] and its distribution and interpretation take place as indicated in sections 3.4.2 and 3.4.2 respectively (note that since PRO keeps its φ - and T-features unvalued, it is still active for further agree).

What happens when PRO* is merged in a derivation with a finite T with which it agrees? Let us assume that Agr in T has a special status in Null Subject Languages (NSL) (see Alexiadou & Anagnostopoulou 1998 for such a proposal) and that this Agr is a pronominal-like element attached to T and already present in the numeration, this is to say that Agr in NSL has similar characteristics as personal pronouns in non NSL.

My proposal is that finite T attracts PRO* due to EPP. Once in [Spec, TP], PRO* agrees with finite T since it has unvalued φ - and T-features and two agree-chains are formed: one to share the valued φ -features and the other one to share the valued (nominative) T feature. By doing this, the interpretable φ features and nominative Case of PRO* get valued, or in other words, PRO becomes *pro*. Furthermore, I would like to propose that different languages might have the option of spelling out this instantiation of PRO*, giving rise to the nominative personal pronominals. It is a strong claim since it follows that personal pronouns in NSL are never in the numeration as such whereas the personal pronouns in non-NSL are in the numeration. In other words, personal pronouns of non-NSL are lexical items present in the numeration whereas personal (nominative) pronouns in NSL are spelled out instantiations of PRO* once its φ -features are valued. In NSL, personal pronouns would be like extra elements for diverse semantic, pragmatic and/or syntactic effects.

This concerns "real" or "complete" NSL like Spanish, Italian or European Portuguese. Partial NSL such as Finnish, Hebrew or Brazilian Portuguese, have null subjects, but these are constrained. Holmberg (2005) argues that these languages do not make use of *pro* but of PF-deletion, i.e. the pronouns are already present in the numeration and their phonological content is deleted at PF interface.

Object *pro* in languages like Spanish and Italian deserves extensive research and discussion, yet this is beyond the scope of this thesis so I will leave this question open for future research.

3.7.2. OC in Hebrew finite clauses

The case of OC in Hebrew finite clauses is extensively discussed by Landau (2004b). Hebrew is a partial NSL, i.e. null subjects are allowed only in past and future tenses, and only in first and second persons but not in third. Nonetheless, 3rd person *pro*-drop in Hebrew is possible in non-referential contexts, i.e. quasi-argument *pro* (like in weather predicates), generic arbitrary *pro* or expletive *pro*. The general agreement is that the defective nature of Hebrew present tense (lacking any person morphology) is crucially involved in not allowing *pro*-drop, as long as one accepts Rizzi's (1986) claim that the feature [person] is the responsible for the identification of *pro*.

Despite this, a null 3rd person pronoun does exist in Hebrew embedded sentences as long as it is bound by a matrix argument, as shown in (25), repeated here as (126).

- (126) Hem_i kivu se-hem_{i/j}/pro_{i/*j} yelxu ha-bayta mukdam. (Landau 2004b:816)
 They hoped that-they/pro will-go_{3rd} home early
 They hope that they would go home early
 "They hoped to go home early"

Landau argues that the null subject in (126) is actually PRO. He argues that 3rd person *pro* does not exist in Hebrew and OC PRO surfaces in these contexts because both I^o and C^o have a [+R] feature, which can check off one another (see section 2.4). Nothing then prevents the presence of PRO, which is [-R].

This kind of OC PRO appears in other languages such as Finnish and Brazilian Portuguese (Holmberg, 2005). He notes that this takes place only in partial NSL and argues that these languages are not really NSL in the sense that they do not resort to *pro* but to PF deletion of the personal pronouns. The null subject in sentences like (126) is neither *pro* nor PRO, but a null 3rd person logophor (in Holmberg's terms). This logophor must be bound by an antecedent, yielding OC. However, it can appear in subject position of matrix clauses, where it cannot be bound, hence its generic interpretation (as a kind of last resort).

In conclusion, OC in finite clauses in languages like Hebrew, Finnish or Brazilian Portuguese, does not pose a problem for the SE-PRO proposal: these languages have a

null 3rd SE-anaphor besides the "standard" null SE-anaphor PRO. As a SE-anaphor, it must form a chain whenever this is possible, namely, when it is in subject position of an embedded sentence. However, if chain formation is not possible, it still may be interpreted with *arb* value, the last resort to interpret a SE-anaphor.

This null 3rd person SE-anaphor is similar to Dutch *zich* or Romance *se/si* but without phonological content. Hence, their different syntactic behaviour follows: the null 3rd SE-anaphor may appear in subject position of embedded finite sentences because it can be licensed by means of long ECM (although it is not at the phonological border, it still may undergo long ECM due to its null phonological content). Dutch *zich* would make similar derivations crash at PF since it does not allow long ECM unless it is at the PB, which is not the case in embedded finite clauses that usually have overt C.

3.7.3. European Portuguese inflected infinitives

Inflected infinitives in Portuguese allow overt subjects and *pro* while PRO is prohibited (see Raposo 1987).

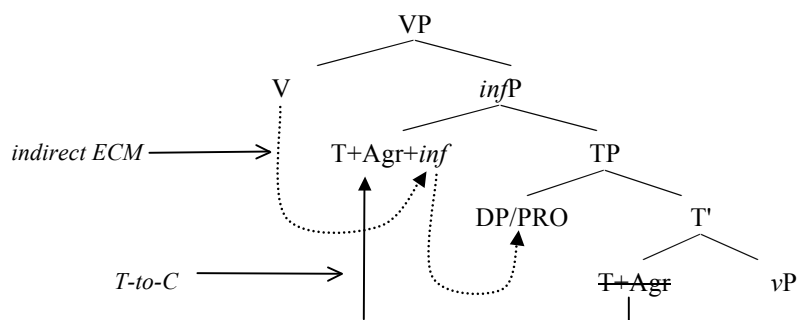
(127) Será difícil [eles / *PRO aprovarem a proposta].

(Raposo 1987:86)

Will be difficult [they / *PRO approve_{infinitive+Agr} the proposal]

"It will be difficult they to-approve-Agr the proposal (...for them to approve...)"

(128)



This can be accounted for by saying that the overt φ -features of the inflected infinitive value PRO (recall from the discussion in 3.7.1 that Portuguese is a NSL so the φ -features in T are present in the numeration and have pronominal characteristics), turning it into *pro*. If we assume that these infinitivals have a similar structure to Spanish infinitivals, as represented in (128), it can be argued that the subject is licensed by

indirect ECM. That explains why A-movement of the subject into the matrix clause is disallowed.

3.7.4. OC into adjuncts

As stated in section 3.2.2.1 adjunct control is one of the phenomena that Hornstein (1999) uses to motivate his MTC approach. The apparent paradox posed by structures such as (129), is essential to his discussion:

(129) a) $NP_1 V NP_2$ [adjunct $PRO_{i/*2} \dots$]

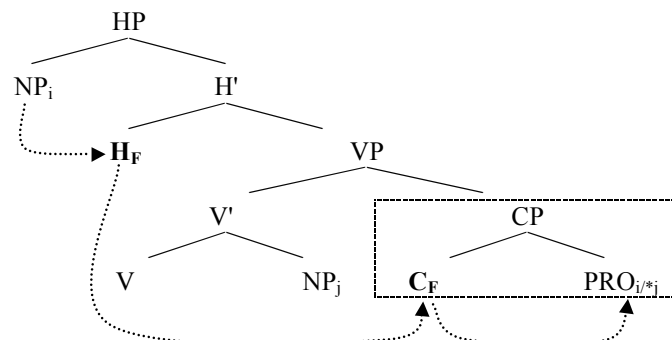
b) $John_i$ reviewed every book $_j$ [without $PRO_{i/*j}$ reading $it_{i/j}$] (Hornstein 1999:88)

On the one hand *every book* c-commands *it*, given the fact that it can bind *it*, on the other hand it should not count as an intervener between PRO and *John*. This is indeed a puzzle if one assumes that adverbials of this type are mere adjuncts, and not explicitly connected to the structure they modify. However, there is considerable evidence discussed by Cinque in a range of works (e.g. Cinque 1999) that this view is too simplistic. According to Cinque, adverbials are directly connected to the functional structure in the extended projection of the verb (to use Grimshaw's (1991, 2000) terms). Cinque assumes that adverbials are generated either as heads or as specifiers of functional heads. Somewhat varying on this proposal, the following assumptions may be made:

- i. the nature of an adverbial clause is encoded in its C-system
- ii. functional heads associated with adverbial functions may probe for and induce agreement with the C-system of the adverbial clause

Given Cinque's structures, the functional head in question will always be higher than the direct object (either in its base position or in its derived position). Thus, omitting much detail, as in (130).

(130)



Assuming, as earlier, that OC is parasitic on independent Agree relations, NP₂ will be bypassed, and only NP₁ will be available. Clearly, this proposal has many ramifications that lead beyond the scope of the present thesis. Hence, exploring its full range of consequences will have to wait for another occasion.

3.8. Conclusions

This chapter has provided an alternative analysis of control. PRO is conceived as a null SE-anaphor, a kind of anaphor for whose existence independent evidence has been provided in the literature (see among others Reinhart & Reuland 1993, 1995; Reuland 2001). Its interpretation is derived from the binding theory as conceived by Reuland (2001) and its distribution follows naturally from its Case requirements and Phase Theory. From this perspective, the Calculus provided by Landau (2004b) is no longer needed in order to account for the distribution of PRO and Control Theory is reduced to binding processes. The ultimate goal is to eliminate both Control Theory and the Theory of *pro* by analyzing the empty category *pro* as an instantiation of PRO in a syntactic environment where T has overt and valued φ -features. Nonetheless, this goal, and in particular what further regards *pro*, will remain open for future research.

Finally, a continuum between PRO and pronominals can be defined as in table (131) below, where other elements like SE-anaphors and the null 3rd logophor described by Holmberg (2005), are to be included. I will provide a more complete picture of this continuum in what rests of the dissertation.

(131) Continuum PRO-Pronominals:

	PRO	PRO*	null 3rd logophor	zich / si	pro	he/him/his
Referential independence	[+R]	[+R]	[-R]	[-R]	[+R]	[+R]
φ-Features	$i\varphi_{\text{person}} [-]$ $i\varphi_{\text{number}} [-]$	$i\varphi_{\text{person}} [-]$ $i\varphi_{\text{gender}} [-]$ $i\varphi_{\text{number}} [-]$	$i\varphi_{\text{person}} [3^{\text{rd}}]$	$i\varphi_{\text{person}} [3^{\text{rd}}]$	$i\varphi_{\text{person}} [\text{val}]$ $i\varphi_{\text{gender}} [\text{val}]$ $i\varphi_{\text{number}} [\text{val}]$	$i\varphi_{\text{person}} [3^{\text{rd}}]$ $i\varphi_{\text{gender}} [\text{♂}]$ $i\varphi_{\text{number}} [\text{sing}]$
Structural Case	uT [val]	uT [val]	uT [val]	uT [val]	uT [val]	uT [val]
Phonological content	NO	NO	NO	YES	NO	YES
Long ECM	YES	YES	YES	NO	YES	NO
Chain formation	YES	NO	YES	YES	NO	NO

Chapter 4

Romance *se/si* clitics

4.1. Introduction

In this chapter I will introduce *se/si* clitics in Romance and their counterparts in Slavic. First I will review the most relevant literature on their status and function. I will furthermore put forward my hypotheses regarding these clitics in the different constructions that will be developed in subsequent chapters, as well as their relation with similar particles in Germanic.

4.1.1. Clitics and cliticization in Romance

The Romance pronominal system distinguishes between tonic and non-tonic (clitic) pronouns, as summarized in table (1) for Spanish.

(1) Spanish pronominal system: *(table based on Fernández Soriano 1993)*

	TONIC	NON-TONIC (CLITICS)		
	nominative	accusative	dative	reflexive
1 st person singular	yo	me	mí	me
2 nd person singular	tú	te	ti	te
3 rd person singular	él, ella, ello	lo, la	le	se
1 st person plural	nosotros	nos	nos	nos
2 nd person plural	vosotros	os	os	os
3 rd person plural	ellos, ellas	los, las	les	se

As Fernández Soriano (1993) points out, the Latin origin of the non-tonic pronouns seems clear. The first and second person clitic pronouns are derived from the corresponding tonic pronouns, and the third person clitic pronouns are derived from the third person demonstrative. Tonic third person pronouns *él, ella, and ello* (it) are derived from Latin nominative third person demonstratives *ille, illa, illud*. Non-tonic or clitic accusative third person pronouns *lo, la, lo* are derived from the accusative forms of the third person demonstrative *illu, illam, illud*. Finally, the dative non-tonic third person pronoun *le* is derived from the dative third person demonstrative *illi*. This shows

that Romance, Spanish in our case, has conserved residues of the Latin case marks in its pronominal system.

Authors such as Strozer (1986) and Fernández Soriano (1989, 1993) have studied clitic pronouns in Spanish. They do not bear accent, unlike morphological independent words. They are phonologically dependent on the verb and cannot appear isolated (2). They cannot be coordinated (3) nor elided (4). They obligatorily adjoin to the verb and nothing (except another clitic) can intervene between the verb and the clitic (5). When two or more clitics appear in the same predicate, they follow a very strict person (6) (cf. Perlmutter 1970) and Case (7) (cf. Dinnsen 1972) order.

- (2) a. ¿Quieres carne o pescado? Carne. (Fernández Soriano 1993:22)¹
 Want meat or fish? Meat
"Do you want meat or fish? Meat."
- b. Lo quieres o la quieres? *Lo. (Fernández Soriano 1993:22)
*It*_{clitic masculine} want or *it*_{clitic feminine} want? **It*_{clitic masculine}.
"Do you want him or do you want her? Him."
- (3) a. Juan trajo **el coche** y **la moto**. (Fernández Soriano 1993:23)
 Juan brought **the car** and **the motorbike**
- b. *Juan **lo** y **la** trajo. (Fernández Soriano 1993:23)
 Juan *it*_{clitic masculine} and *it*_{clitic feminine} brought
- (4) a. Juan trajo y llevó **el coche**. (Fernández Soriano 1993:23)
 Juan brought and took **the car**
- b. Juan **lo** trajo y llevó. (Fernández Soriano 1993:23)
 Juan *it*_{clitic} brought and took
"Juan brought it and took it."
- (5) a. *Lo no quiero. (Fernández Soriano 1993:23)
*It*_{clitic} not want
"I don't want it."
- b. No lo quiero.
 Not *it*_{clitic} want
"I don't want it."

¹ The glosses and translations of all the examples from Fernández Soriano (1993) are mine.

- (6) SE II I III (II=2^a p., I=1^a p., III=3^a p.) (Fernández Soriano 1993:24)
- a. Se [me/te/le] cae / se lo compró. (Fernández Soriano 1993:24)
 Se [me/te/le] falls / se lo bought
- b. Te [me/nos] fuiste / te lo(s) compraste. (Fernández Soriano 1993:24)
 Te [me/nos] left / te lo(s) bought
- c. [Me/te] lo dieron. (Fernández Soriano 1993:24)
 [Me/te] lo gave
- (7) REFLEXIVE BENEFACTIVE DAT. ACC. (Fernández Soriano 1993:24)
- a. *Te me escapé (te me escapaste). (Fernández Soriano 1993:24)
 *Te me escaped (te me escaped)
- b. *Me le acerqué (me lo acerqué). (Fernández Soriano 1993:24)
 *Me le got closer (me lo got closer)

Not only do third person clitic pronouns distinguish accusative and dative cases but table (1) shows that they also have a differentiated reflexive form: *se*. This chapter will focus precisely on this reflexive clitic from the next section onwards. Since *se* (*si* in other Romance languages like Italian) is a clitic that appears associated not only to reflexive construction but to many other (ergatives, middles, inherent reflexives or ergatives), I will call it *se/si* clitic. In this chapter, I will focus on the accusative and dative non-reflexive object clitics.

The first works on clitics within generative grammar, such as Kayne (1975) for French and Italian, claimed that they were independent syntactic elements. *Clitic climbing* was a piece of evidence that supported this thesis: a clitic could optionally raise (move) from an embedded to a matrix clause as in (8). Hence, cliticization hypothesis: clitics were assumed to be base-generated in canonical object positions, i.e. A-positions, and they adjoined to the verb, i.e. to an A'-position (Kayne 1975, 1991).

- (8) a. [Quiero / puedo] dár**selo**. (Fernández Soriano 1993:43)
 [Want / can] give-**se**_{dative clitic}-**lo**_{accusative clitic}
- b. **Se** **lo** [puedo / quiero] dar. (Fernández Soriano 1993:43)
Se_{dative clitic} **lo**_{accusative clitic} [can / want] give
 "I can/want give it to him/her."

Clitic pronouns are, on the other hand, very similar to the verbal inflection. In fact, many authors (Mendikoetxea 1992, Borer 1985, among others) argue that they are the spell-out of certain features of the verbal morphology (agreement, Case, etc.). One

crucial piece of evidence for this hypothesis on the nature of the clitics is the so-called *clitic doubling*. This consists of doubling an overt argument (direct object or indirect object) by means of a clitic as in (9). The underlying hypothesis is that clitics are elements base-generated in the position where they appear (adjoin to the verb, i.e. an A'-position).

(9) a. **Me** lo han dicho **a mí**. (Fernández Soriano 1993:30)

Me_{dative clitic} lo_{accusative clitic} have said **to me**

"They have said it to me."

b. Se lo han dicho a Juan. (Fernández Soriano 1993:30)

Se_{dative clitic} lo_{accusative clitic} have said **to Juan**

"They have said it to Juan."

Other authors provide alternative analyses like Uriagereka (1992) and Torrego (2008a,b), who argue that clitics are determiners that license a *pro* in complement position. Sportiche (1992), on the other hand, claims that clitics head a clitic voice phrase in whose specifier position is the null or overt argument in order to be licensed by checking any feature.

We see that clitics seem to be both independent syntactic elements, on the one hand, and verbal affixes, on the other hand. I will explain later on (in section 4.7.4) that my own hypothesis on *se/si* clitics is that they are independent syntactic elements base-generated in A-positions. More concretely, I will claim that *se/si* clitics are ϕ -defective SE-anaphors base-generated in A-positions (either subject or object). Their non-tonic nature comes from their defectiveness on ϕ -features, which Romance languages mark on the phonological component.

4.1.2. Romance *se/si* clitics

Romance *se/si* clitics are attested in several contexts across Romance languages. Here I will provide examples in Spanish, Italian, French, and Portuguese.

First, they are usually required in reflexive (10) and reciprocal contexts (11), either with or without morphologically complex anaphors like *sí mismo* (himself) in (10a)² and *el uno al otro* (each other) in (11a).

² In chapter 5 I will argue that this optionality is apparent rather than real because in some cases the complex anaphor has to appear, and in other cases its presence triggers semantic differences.

- (10) a. María ***(se)** critica. (Spanish; Torrego 1995:223)
 María **se** self-criticizes
"María criticizes herself"
- b. Maria ***(si)** guarda. (Italian; Burzio 1986:37)
 Maria **si** watches
"Maria watches herself"
- c. Luc **se** lave. (French; Labelle 2008:834)
 Luc **se** wash_{present.3s}
"Luc is washing (himself)."
- (11) a. ***(Se)** miran los unos a los otros. (Spanish; Torrego 1995:224)
(Se) self-look the ones to the others
"They look at each other"
- b. I bambini **si** parlano. (Italian; Manzini 1986:248)
 The children **si** talked
"The children talk to each other."
- c. Os meninos insultaram-**se**. (Portuguese; R&U 1996:801)
 The children insulted-**se**
"The children insulted themselves (each other)."
- d. Luc et Pierre **se** regardent. (French; Labelle 2008:834)
 Luc and Pierre **se** look-at_{present.3p}
"Luc and Pierre look at each other."

They are also present in inchoative or ergative constructions (12) and inherent ergative or reflexive constructions (13)³.

- (12) a. El cristal **se** rompió. (Spanish; Mendikoetxea 2008:291)
 The glass **se** broke
"The glass broke."
- b. Il veto **si** rompe. (Italian; Burzio 1986:37)
 The glass **si** breaks
"The glass breaks (itself)"

³ The difference between inherent ergative verbs and ergative verbs is that only the latter have a transitive (causative) alternation as shown in (i):

- (i) a. The glass broke. (inchoative / ergative / unaccusative alternate)
 b. John broke the glass. (causative / transitive alternate)

c. La branche **s'est** cassée. (French; Dobrovie-Sorin 2006:121)

The branch **se**-has broken

"The branch broke."

(13) a. Juan **se** arrepintió de lo que dijo. (Spanish)

Juan **se** regretted of that_{article} that_{comp} said_{3rd,sing}

"Juan regretted what he said."

b. Giovanni **si** sbaglia. (Italian; Burzio 1986:37)

Giovanni **si** mistakes

"Giovanni mistakes (himself)"

c. Marie **s'est** souvenu de Jean. (French; Dobrovie-Sorin 2006:169; endnote 10)

Marie **se**-has remembered of Jean

"Marie remembered Jean."

These examples of *se* (reflexive, reciprocal, ergative and inherent ergative) can be inflected for person and number, as can be seen in (14) below for Spanish. Burzio (1986) called them *se/si* (as opposite of SE/SI, as we will see below).

(14) a. Juan **se** lava. d. Juan y María **se** lavan.

Juan **se** washes

Juan and María **se** wash

b. Yo **me** lavo. e. Nosotros **nos** lavamos.

I **me** wash

We **nos** wash

c. Tú **te** lavas. f. Vosotros **os** laváis.

You **te** washes

You_{plural} **os** wash

Also these clitics can be found in a wide variety of arbitrary constructions, as in impersonals (15), passives (16), and middles (17).

(15) a. Aquí **se** lee libros. (Spanish; Mendikoetxea 2008:316)

Here **se** read_{singular} books

"One (SE) reads books (here)."

b. **Si** leggerà_{singular} volentieri alcuni articoli_{plural}. (Italian; Burzio 1986:43)

Si will read_{sing} willingly a few articles_{plural}

"We will be eager to read a few articles."

c. Compra-**se** sempre demasiadas salsichas no talho Sanzot. (R&U 1996:750)

Buy_{singular}-**se** always too many sausages at the butcher shop Sanzot

"One (people) always buy too many sausages at the Sanzot butcher shop."

- (16) a. **Se** comen_{plural} las manzanas_{plural} (Spanish; Mendikoetxea 2008:291)
Se eat_{plural} the apples_{plural}
"The apples are eaten."
- b. Le materie letterarie_{plural} **si** studiano_{plural} in questa università. (Belletti 1982:1)
The humanities_{plural} **si** study_{plural} in this university
"Humanities are studied in this university."
- c. Essas salsichas compraram-**se** ontem no talho Sanzo. (R&U 1996:750)
Those sausages_{plural} bought_{plural}-**se** yesterday in the butcher-shop Sanzot
"Yesterday someone or other bought those sausages at the Sanzot butcher."
- d. Les pommes **se** mangent en hiver. (Dobrovie-Sorin 2006:122)
Apples **se** eat during winter
"Apples are eaten during winter."
- (17) a. Estas manchas no **se** quitan con nada. (Mendikoetxea 1999:1635)⁴
These stains not **se** come out with nothing
"These stains don't come out at all."
- b. Le grec **se** traduit facilement. (Dobrovie-Sorin 2006:122)
the Greek **se** translates easily
"Greek translates easily."
- c. Questo vestito si lava facilmente. (Cinque 1988:559)
This suit **si** washes easily
"This suit washes easily."

These instances of *se/si* are invariable, i.e. they are never inflected for person or number. Burzio distinguish them from *se/si* by calling them SE/SI⁵.

The impersonal subject of the impersonal and passive *se* constructions is interpreted as an arbitrary human subject. Cinque (1988) developed the Theory of Arb in order to study the interpretation of the subject of these constructions among others. Chierchia (1995) and Mendikoetxea (2002, 2008) also have contributed to the semantics of arbitrary subjects.

In Spanish, the study of these clitics can be tracked back to the works of Strozer (1976) and Martin Zorraquino (1979). Sánchez López (2002) gives an extensive

⁴ Glosses and translation are mine.

⁵ Burzio did not actually talk about middle *si* in Italian, but I will include this construction due to its impossibility to be inflected for neither person nor number.

overview of the works on this clitic. Also in the framework of Cognitive Grammar the Romance clitics have been studied. Maldonado (1999) uses the term *middle voice* to define the aforementioned uses of the clitic *se* in Romance languages (and others), and recognizes in it a system of middle voice, which defines as a basic voice system (not derived like the passive voice) that corresponds to a different alternative of conceptualization of the event denoted by the predicate. In Reference & Role Grammar works on Romance clitics can be found such as González Vergara (2006).

Se/si clitics are also attested in Slavic languages as Rivero (2001) shows⁶ for Polish. As in Romance, these clitics are present both in reflexive contexts and in non-reflexive ones: middle constructions, ergative verbs, inherent ergative verbs and impersonal constructions. I will discuss this data in section 4.7.2.

The question as to whether Germanic languages have counterparts of Romance and Slavic *se/si* clitics will be postponed until the end of the chapter in section 4.7.3.

4.1.3. Constructions with *se* in Spanish

Mendikoetxea (1999) provides a descriptive classification of the different variants of arbitrary *se* construction in Spanish: middle (18d), passive (18a) and impersonal (18b,c)⁶.

- (18) a. **Se** pasaron los trabajos a ordenador. (Mendikoetxea 1999:1635)
 Se wrote the papers to computer
"The papers were typed in the computer."
- b. **Se** agasajó a los invitados. (Mendikoetxea 1999:1635)
 Se entertained to the guests
"The guests were entertained."
- c. Por aquí **se** llega antes a Madrid. (Mendikoetxea 1999:1635)
 By here **se** arrive earlier to Madrid
"This way one comes earlier to Madrid."
- d. Estas manchas no **se** quitan fácilmente. (Mendikoetxea 1999:1635)
 These stains not **se** come out easily
"These stains don't come out easily."

⁶ Glosses and translation of examples from Mendikoetxea (1999) are mine.

The clitic *se* can also be found in Spanish inchoative, reflexive and pseudoreflexive (inherent ergative) constructions with *se* in (19).

- (19) a. El cristal **se** rompió. (Mendikoetxea 2008:291)
 The glass **se** broke
"The vase broke."
- b. María *(**se**) critica. (Torrego 1995:223)
 María **se** self-criticizes
"María criticizes herself"
- c. Juan **se** arrepintió.
 Juan **se** changed-his-mind
"Juan changed his mind."

Finally, we find in Spanish other constructions where *se* appears, and where it seems to have an aspectual contribution (*aspectual se*) or to introduce an extra argument that is not selected by the verb (*dative se*).

- (20) a. Juan **se** comió las manzanas. (Otero 1986:87)
 Juan **se** ate the apples
"Juan ate up the apples."
- b. Este niño no **me** come nada.
 This child not *me*_{dative} eats nothing
"This child eats nothing (and it affects me)."

Mendikoetxea (1999) proposes a classification of the constructions with impersonal *se* in (18) in two groups, having each the possibility of presenting a middle reading⁷:

- 1) **Passive *se*** (18a): these sentences do not present any aspectual, semantic or thematic restrictions. The subject is implicit, generally not delimited, and not recoverable unlike periphrastic passives (this means that it cannot be made explicit). The unmarked position of the subject is post-verbal. Verbs that select cognate objects are allowed in this construction and it has a less intentional reading than the periphrastic passive. There is verb-object agreement (21a).

⁷ Note however, that Mendikoetxea does not argue that there are different *se* clitics. As a matter of fact, the author points out that her work deals with one and the same lexical element. The distinction is established for descriptive purposes.

Nonetheless, there also exists a *non-standard expression*⁸ as in (21b) where there is no verb-object agreement.

- (21) a. **Se venden libros.** (Mendikoetxea 1999:1676)
Se sell_{plural} books
"The houses are sold"
- b. **Se vende libros.** (Mendikoetxea 1999:1676)
Se sell_{singular} libros
"One sells books."

- a) **Middle-passive *se*** (18d): the subject cannot be expressed (unlike the passives with *se*), and the canonical position of the object is pre-verbal (unlike the passives with *se*)⁹. The subject is determined (unlike the passives with *se*) and this construction is restricted to agentive subjects of transitive verbs. The subject is an inanimate entity (singular or plural). This construction needs linguistic elements that activate the genericity. The interpretation is stative. Finally, the predication is over a property or class rather than over an individual.
- 2) **Impersonal *se*** (18b): in this construction the notional object is the grammatical object. It happens with verbs that select prepositional objects, clauses and small clauses, as well as with verbs that are non-transitive: intransitive, unaccusatives, copulatives, and periphrastic passives. This construction shows aspectual and lexical restrictions.
- a) **Middle-impersonal *se*** (18c): the canonical subject position is pre-verbal and the subject is not to be specific (unlike middle-passives with *se*). This construction is not thematically restricted, and it admits experiencer subjects. Unlike middle-passives with *se*, this construction also admits inanimate entities as well as a singular animate entity if a quality of such entity is predicated. Activators of the genericity are necessary. In this case, the predication is also over a property or class, rather than over an individual.

⁸ Mendikoetxea's original term is *giro no concertado* (Mendikoetxea 1999:1676). The translation is by the author (IITG). Mendikoetxea (1999:1676-1680) points out that this construction (21b) is marginal.

⁹ Note that the subject in passive *se* can also be pre-verbal. The difference between passive *se* and middle-passive *se* is that the subject of middle passive *se* is a topic, according to Mendikoetxea (1999).

The implicit subject is semantically more present in the periphrastic passive construction. In passives and impersonals with *se*, the subject is semantically less present than in periphrastic passives, though more present than in middles (middle-passives and middle-impersonals). In turn, the subject is the least present in inchoative constructions (even less than in middles).

4.2. Impersonal and passive SE

4.2.1. Literature review

Belletti (1982) studies in her seminal work the Italian impersonal and passive constructions with the clitic *si*, as illustrated in the examples (23) and (22) respectively:

(22) a. I **dolci**_{plural} al cioccolato **si mangiano**_{plural} in questa pasticceria. (Belletti 1982:1)
 cookies chocolate **si** eat_{plural} in this pastry shop
"Chocolate cookies eat (si) in this pastry shop."

(23) a. In questa pasticceria **si mangia**_{singular} soltanto i **dolci**_{plural} al cioccolato. (Belletti 1982:1)
 In this pastry shop **si** eat_{singular} only cookies chocolate
"In this pastry shop one (si) eats only chocolate cookies."

b. In questa universitá **si studia**_{sing} le materie **letterarie**_{plur.}. (Belletti 1982:1)
"In this university one (si) studies humanities."

c. **Si va**_{sing} al cinema un po' troppo di rado ultimamente. (Belletti 1982:1)
"One (si) goes to the movie too rarely recently."

d. No **si dorme**_{sing} piú con tutto questo rumore. (Belletti 1982:2)
"One (si) does not sleep anymore with all this noise."

Belletti states that sentences in (22) are instances of the **morphological passive or passive *si*** construction in Italian. This is another form to passivize in Italian by means of the clitic *si*, which has the same functions as the passive morphology, i.e. it absorbs the theta role for the nominal in subject position and it absorbs the accusative Case for the object position. Thus, the clitic *si* is a clitic pronoun base-generated under the node INFL. It is a pronominal lexical item that has to be Case marked and assigned a θ -role. If *si* is generated in INFL and thence, INFL is pronominal, INFL has Case (nominative) and it bears the θ -role that would be otherwise assigned by VP to the subject NP. The clitic also absorbs the accusative Case that would be otherwise assigned by V to its

direct object NP. Note that morphological passives with *si* in Italian allow post-verbal subjects (24), as expected. They do not allow object clitics (25) since no accusative Case is available. Nor do they allow *da*-phrases as in (26), unlike in standard passives as in (27), since the clitic *si* retains the external theta role unlike passive morphology.

- (24) a. **I dolci al cioccolato** *si* mangiano in questa pasticceria. (Belletti 1982:6)
 b. *Si* mangiano **i dolci al cioccolato** in questa pasticceria. (Belletti 1982:6)
"(Si) eat chocolate cookies in this pastry shop."
- (25) a. ***Li** *si* mangiano in questa pasticceria. (I dolci al cioccolato) (Belletti 1982:7)
"Them_{clitic} (si) eat in this pastry shop."
 b. ***Le** *si* studiano in questa università. (Le materie letterarie) (Belletti 1982:7)
"Them_{clitic} (si) study in this university."
- (26) a. *I dolci al cioccolato **si** mangiano in questa pasticceria **da Mario**. (Belletti 1982:7)
"Chocolate cookies (si) eat in this pastry shop by Mario."
 b. *Le materie letterarie **si** studiano in questa università **da molti studenti**. (Belletti 1982:7)
"Humanities (si) study in this university by many students."
- (27) a. I dolci al cioccolato **sono stati** mangiati **da Mario**. (Belletti 1982:7)
"Chocolate cookies have been eaten by Mario."
 b. Le materie letterarie **sono studiate da molti studenti**. (Belletti 1982:7)
"Humanities are studied by many students."

Sentences in (23) are instances of what Belletti calls **impersonal *si*** construction. It is characterized by the fact that the verb does not agree with the object in number, and it always shows up with the features third person singular. Moreover, this construction allows virtually any kind of verb: transitives (23a,b), unaccusatives (23c) and unergatives (23d). Belletti analyses this construction as an instance of the Pro-drop parameter. In (23) *si* is an INFL clitic that is assigned (absorbs) nominative Case rather than accusative as in (22). Hence, the direct object in [V, NP] position can be assigned accusative Case. Nevertheless, *si* bears the external θ -role since INFL is pronominal (due to the Pro-drop parameter setting in Italian). In this case, INFL is a proper governor of an empty category in subject position [NP, S] because it is pronominal and is coindexed with such empty category. Therefore, this category satisfies the Empty Category Principle (ECP). These sentences do not allow overt subjects because that

would violate the Theta Criterion and the Case Filter. Moreover, object clitics are allowed because accusative Case is available (28).

(28) a. (I dolci al cioccolato) Li si mangia volentieri in questa pasticceria.

(Belletti 1982:13)

"(As for chocolate cookies) One eats them with pleasure in this pastry shop."

Finally, Belletti notes that French has passive *si* constructions (29) similar to their Italian counterparts (22), while it lacks impersonal *si* constructions (30) unlike Italian (23).

(29) a. Les noîsettes **se** mangent.

(Belletti 1982:19)

The hazelnuts **se** eat_{plural}

"The hazelnuts are eaten."

(30) a. ***Se** mange les noîsettes.

(Belletti 1982:20)

***Se** eat_{singular} the hazelnuts

"One eats hazelnuts."

This is expected since French is not a Pro-drop language and hence it does not allow a pronominal INFL that properly governs a null subject (violating the ECP). I will come back to French later on in this section.

Burzio (1986) attempts to analyse the clitic *si* in all the possible constructions in which it can appear in Italian. On the one hand, he makes a first distinction between *si*, which can be found with reflexive, ergative and inherent reflexive verbs, and SI, which can be found in the impersonal and passive constructions in (23) and (22) above. For the time being, I will focus on Burzio's analysis of SI, i.e. of the clitic *si* in passive and impersonal constructions, while I will put apart the rest of cases of *si* (reflexive, ergative and inherent reflexive) until the following sections. Burzio considers impersonal SI a subject clitic that forms a chain with an empty category *ec* in subject position. The chain [*ec, si*] bears the subject θ -role as well as nominative Case. Hence, SI cannot appear in environments where no Case is available like infinitivals in (31)¹⁰.

(31) E' necessario teleforane/*telefonarsi a Giovanni.

(Burzio 1986:44)

It is necessary to phone/**si**-to phone to Giovanni

¹⁰ Though see Cinque (1988) below, who discusses the availability of SI in untensed contexts where nominative is available, i.e. Aux-to-Comp and raising constructions.

SI can be a derived subject and hence it is allowed with passives as in (32). This supports the claim that SI cannot be base-generated in subject position. If the *ec*-SI chain were base-generated, a violation of the Projection Principle would ensue since SI would fail to receive the internal θ -role at D-Structure

- (32) a. **Si** è stati invitati. (Burzio 1986:44)
Si is been invited
"We have been invited."

Burzio assumes that SI is a clitic restricted to subject position in the sense that it needs to be in subject position in order to cliticize, though it can be base-generated in any NP position. The solution to (32) is based on movement: SI is inserted in object position at D-structure, moving to subject position and afterwards cliticizing from that position.

As for passive SI (when the verb agrees with the object), Burzio claims that, despite the fact that the object is clearly in subject position (pre-verbal position and agreement with the verb), it has been base-generated in object position (complement of the verb) and thence, moved to subject position, whence it triggers agreement with the verb.

Manzini (1986) tries to provide a unified analysis of all instances of *si* in Italian, as Burzio did. She proposes one single lexical item *si*, from which four types are obtained: impersonal, reflexive, middle and middle-reflexive *si*. As I did for Burzio's analysis, I will focus on Manzini's analysis of impersonal *si*, whereas I will put aside the other instances of *si* until following sections.

Given that impersonal *si* and PRO_{arb} have a similar interpretation¹¹, Manzini argues that the interpretation of both elements is due to the fact that they introduce a free variable. As can be seen in (33) and (34), the verb can show singular agreement whereas the adjective shows plural agreement morphemes. This leads Manzini to claim that *si* has a third person feature along with number and gender features unspecified. She claims that impersonal *si* is an argument subject to the Theta Criterion, and always associated to subject θ -position. It has an N categorial feature, which makes it to be subject to the Case Filter and associated with (exclusively) nominative Case. It is thought to be a clitic on a verb.

¹¹ See especially Hernanz (1994) for this issue.

- (33) Si **e'**_{singular} facilmente **nervosi**_{plural,masculine}. (Manzini 1986:242)
 Si **is**_{singular} easily **nervous**_{plural,masculine}
 "One is easily nervous."

- (34) Si **é**_{singular}' **invitati**_{plural,masculine} volentieri. (Manzini 1986:243)
 Si **is**_{singular} **invited**_{plural,masculine} gladly

The subject position of the sentences with impersonal *si* must be filled with an expletive, which in the case of Italian this can be an empty category *ec*. The object position in sentences with derived subjects, like with unaccusative verbs, is filled with a trace left by an empty category *ec* that has undergone A-movement to subject position.

To conclude Manzini's analysis, she argues that the lexical entry for impersonal *si* can be summarized in (35):

- (35) Impersonal *si*: (Manzini 1986:247)
- a. Variable (free/dependent).
 - b. Argument.
 - c. Categorical feature: N.
 - d. Φ -features: third person, unspecified number and gender.
 - e. Clitic on a verb.
 - f. Bound to its subject.

Otero (1986), like Chomsky, Belletti and Burzio, distinguishes two kinds of *se* in impersonal constructions in Spanish: SE-active in (36a) and *se*-passive in (36b).

- (36) a. [*ec se* comió las manzanas] (Otero 1986:87)
ec se eat_{singular} the apples
- b. [Δ *se* comen las manzanas] (Otero 1986:87)
 Δ *se* eat_{plural} the apples

For the *se*-passive construction in (36b) Otero follows Chomsky (1981) and Belletti (1982): *se* absorbs the external θ -role, as well as the accusative Case, in a parallel way as passive morphology. However, he rejects Chomsky's claim of PRO_{arb} as the subject of SE-active in (36a).

Otero argues that *ec* in (36a) has an arbitrary interpretation like PRO_{arb} (it is an *ec**, in Otero's terms). However, this *ec** cannot be PRO because it is never open to control, as can be seen in the contrast between (37) and (38).

- (37) *ec** **se** llegó a saber PRO estudiar (***se**) todos los días. (Otero 1986:91)
 ARB **se** came to know PRO study_{infinitive}(***se**) every day
- (38) *ec** **se** insistió en que *ec** **se** estudia todos los días. (Otero 1986:91)
 ARB **se** insisted in that ARB **se** studies_{indicative} every day

The hypothesis is that *se* is simply a marker of INFL whose functional role is to absorb the definiteness of a finite INFL, encoded in the [\pm def] feature. Finite INFL is [+def] and thus *pro* and other overt pronominals like *él* (he) or *ella* (she), which are [+def] too, are allowed. PRO_{arb} is [-def] and hence it is incompatible with finite INFL. However, when *se* is present, it absorbs the [+def] feature of INFL and renders it as [-def]. In spite of this, PRO_{arb} cannot be inserted because it needs to be ungoverned (see the Theorem of PRO in chapter 2). The only option left is that *ec** is *pro** (*pro_{arb}*), a free pronominal always interpreted as an indefinite human¹².

Otero concludes that *se/si* is an *absorber clitic* in contrast to object clitics (accusative *lo/la*; dative *le*; reflexive *se*) as in (39), which are identifier clitics, i.e. they are necessary in order to identify an empty category in object position, which is *pro* and has a [+def] feature.

- (39) Ana **la** vio.
 Ana **her**_{clitic} saw
 "Ana saw her."

Cinque (1988) found some unexpected asymmetries in the distribution of the impersonal *si* construction in tensed and untensed contexts in Italian.

On the one hand, whereas in tensed contexts this construction is allowed with all major verb classes in Italian (transitive, unergative and ergative intransitive, psych-movement, copulative, passive and raising verbs), it is uniformly excluded in untensed contexts such as untensed control clauses due to a violation of the Case Filter: *si* is a clitic nominal element¹³ that must be part of a CHAIN assigned Case, and the preverbal subject position of the embedded clauses fails to be assigned one.

On the other hand, one would expect then that impersonal *si* be acceptable in untensed clauses in which the preverbal subject position can be assigned Case in some special way. Two such untensed environments are infinitival complements to raising

¹² Although see Mendikoetxea (2002).

¹³ See the analysis of Burzio's (1986) above.

verbs (Chomsky 1981) and untensed clauses involving Aux-to-Comp (Rizzi 1982)¹⁴. This prediction is partially fulfilled: the construction is grammatical in such contexts with transitives and unergative intransitives although it is ungrammatical with all other classes. This asymmetry between transitives and unergatives, on the one hand, and all the other classes, on the other hand, is unexpected under the standard analysis since nominative Case is available for the [NP, IP] position.

The hypothesis that Cinque puts forward states that there are two kinds of impersonal *si*, or rather, *si* can take one out of two values for a determined parameter: its argumenthood. The feature composition of impersonal *si* in (40) implies that it is a clitic pronoun coindexed with the [NP,IP] position (and with Agr when present) that can be an argument or not depending on the value of the feature [\pm argument].

(40) Feature composition of impersonal *si*: (Cinque 1988:530)

- a. syntactically
 - i. [\pm argument]
 - ii. clitic pronoun coindexed with [NP,IP]¹⁵
- b. morphologically
 - i. person: unspecified (generic or arbitrary)
 - ii. number: plural
 - iii. gender: masculine
- c. semantically
 - i. [+human]

When *si* is [+arg], it needs to be associated with some θ -role at every level of representation (D-Structure, S-Structure and LF). Hence, it is only possible with verbs that assign an external θ -role. It dethematizes the position [Spec, IP] (like passives) and retains the absorbed θ -role. Therefore, the θ -role is not recoverable by *da*-phrases¹⁶. In conclusion, [+arg] *si* requires nominative Case and a thematic role, hence it only appears with transitive or unergative verbs.

¹⁴ Aux-to-Comp is allowed at a peculiar stylistic level in the infinitival complement of "verbs of thinking", in adjunct gerundive clauses, in "nominalised" infinitives, and in the infinitival complement of certain nonsubcategorized prepositions (Cinque 1988:524).

¹⁵ So is it coindexed with Agr by transitivity, when it is present.

¹⁶ See Belletti (1982) above.

On the other hand, [-arg] *si* need not be associated with any θ -role at any level of representation. Its function is to provide personal Agr with the features able to identify the content of *pro* as an unspecified generic person pronominal¹⁷. To sum up, [-arg] *si* requires personal Agr and hence, tensed contexts with nominative and personal Agr (this excludes raising and Aux-to-Comp).

(41) Distribution of impersonal *si*:

(table based on data from Cinque 1988)

	Tensed contexts	Untensed contexts	
	Finite clauses	Control clauses	Raising Aux-to-Comp
Transitives	[± argument] <i>si</i>	x	[+ argument] <i>si</i>
Unergatives	[± argument] <i>si</i>	x	[+ argument] <i>si</i>
Ergatives	[- argument] <i>si</i>	x	x
Psych-movement	[- argument] <i>si</i>	x	x
Copulative verbs	[- argument] <i>si</i>	x	x
Passive verbs	[- argument] <i>si</i>	x	x
Raising verbs	[- argument] <i>si</i>	x	x

In tensed contexts [-arg] *si* is always possible since there is personal Agr. Only with transitives and unergatives can appear [+arg] *si*, giving rise to a theoretical ambiguity. In control clauses [+arg] *si* is excluded due to a Case Filter violation (there is no Case available for subject position). In such cases [-arg] *si* is excluded too because there is no personal Agr. In the special untensed contexts that allow nominative subjects, only [+arg] *si* is possible. These environments do not provide personal Agr despite the fact that nominative is available. Therefore, [-arg] *si* is uniformly excluded. However, [+arg] *si* is possible with transitives and unergatives because it can be associated with an external θ -role, and with nominative Case. In the other classes of verbs, [+arg] *si* is excluded, as it happens in tensed contexts.

Si in transitive contexts without verb-object agreement is always [-arg] *si*. The external θ -role is not absorbed but assigned to the generic *pro* in subject position (as it

¹⁷ This idea is similar to the analysis of Otero (1986) above.

happens in intransitive contexts). The verb is then able to assign accusative to its object, and hence the lack of agreement. This is barred from untensed contexts, included raising and Aux-to-Comp. Cinque calls this **impersonal *si***.

Si in transitive contexts where the verb agrees with its object, can be either [+arg] or [-arg]. If *si* is [+arg], it absorbs the external theta role and the verb is no longer able to assign accusative (similar to passive morphology). Hence, the object receives nominative Case, and the verb agrees with it. It is possible in untensed contexts where nominative is available, i.e. raising and Aux-to-Comp. With a specific time reference it implies that an agent took part in the event, so it is compatible with agentive adverbs. Finally, it can control the PRO of a purpose clause, as well as the subject of a small clause. Cinque calls this **impersonal passive *si***.

Mendikoetxea (1992) is concerned with impersonal clitic *se* in Spanish impersonal *se* constructions, which she calls ARB(itrary) SE: Mendikoetxea claims that ARB SE is a functional head. More concretely, it is the spell-out of the person feature of an Agr projection (either AgrS or AgrO), and it absorbs the Case but not the θ -role. In other words, SE is not an argument in the sense of Burzio's work, in which *si* was a clitic base-generated in any NP position and associated with the subject position whence it cliticized.

With unergative and unaccusative verbs, ARB SE is the realization of AgrS. It absorbs nominative Case along with the Agr features of AgrS. Hence, AgrS is no longer a governor and the null element in subject position is PRO. With transitive verbs, the verb can show up invariably in third person singular with ARB SE. In this case, ARB SE is the realization of AgrS and it absorbs nominative and the person feature of AgrS, which no longer is a proper governor. Hence, PRO is inserted in subject position and the object receives accusative case.

The other option is that the transitive verb agrees with the object in constructions with ARB SE. The ARB SE in this case is the realization of the person feature of AgrO rather than of AgrS, so that it absorbs accusative Case. AgrS retains its Agr features and then it is a proper governor able to assign nominative. Since ARB SE has absorbed the accusative Case, the object has to receive nominative from AgrS, and no PRO can be inserted. When ARB SE absorbs accusative Case, it also absorbs the external θ -role (as Belletti claims for *impersonal si* constructions in Italian above) and it assimilates to the passive morphology.

Mendikoetxea (2008) provides a more recent analysis of the impersonal *se* construction in Spanish in particular, and in Romance in general. She calls this kind of constructions *Romance Clitic Impersonal Constructions* (CL-ICs). Mendikoetxea claims that the clitic *si/se* is a pronominal element (Mendikoetxea 2008:295) that lacks number and gender features. However, it is specified for 0-person feature. It is an AGR-clitic¹⁸ based-generated in a position above TP.

The hypothesis that Mendikoetxea puts forward is that *se/si* is an AGR-clitic specified for 0-person feature (though it lacks a categorial D feature, which prevents it to check the EPP feature of T), which checks the person feature of T. As a consequence, it turns a referential T into a non-referential T (similar to Otero's (1986) analysis, Cinque's (1988) [-arg] *si* and to Dobrovie-Sorin's (1998) ACC-*si*).

Furthermore, she argues that the default morphological verbal markers are third person singular and that *se/si* alters the nominative Case assignment by virtue of *se* checking the 0-person feature of T. This is corroborated by the fact that *se/si* is incompatible with a DP that checks nominative Case.

In non-transitive contexts, *se/si* checks the person feature of T but since it lacks categorial D feature, another element is needed to check the D feature of T (EPP). This is done by an empty category (*ec*) in [Spec, ν P]. This *ec* cannot be *pro* because there is no nominative available. Mendikoetxea argues that this *ec* is a special *pro* with number feature but without person feature (in line with some generic pronouns in Holmberg 2005). She calls this instance of *pro* G(eneric)-*pro*, which is possible when T lacks referential person feature¹⁹, which appears with [-finite] T and with [+finite] T with *se/si*. G-*pro* checks the EPP, i.e. the D feature of T (it has no ϕ -features) and *se/si* checks the person feature of T.

In transitive contexts, the agreement of T is "divided": T agrees and checks its person feature with *se/si* whereas it checks its number feature with the agreeing DP in object position. In this construction, ν lacks Case and has deficient Agr features, which checks against G-*pro* in [Spec, ν P] position.

As for the variant where there is no verb-object agreement, Mendikoetxea argues that this is due to the presence of a Loc(ative) argument merge in [Spec, TP] that

¹⁸ Mendikoetxea distinguishes AGR-clitics, which have verbal properties and are present in impersonal clitic construction, from D-clitics, which identify an empty category in object position.

¹⁹ Similar to Otero's (1986) *pro**.

triggers default morphology in the verb and checks the D feature of T (EPP) (Mendikoetxea 2008:316).

Raposo & Uriagereka (1996) study the impersonal *se* construction in European Portuguese, which can be found with two variants: in (42) the verb agrees with the DP object whereas in (43) there is no such agreement.

(42) Ontem compraram-**se** demasiadas salsichas no talho Sanzot. (R&U 1996:750)
 Yesterday bought_{plur}-**se** too many sausages at the butcher-shop Sanzot
"Yesterday someone or other bought too many sasages at the Sanzot butcher shop."

(43) Compra-**se** sempre demasiadas salsichas no talho Sanzot. (R&U 1996:750)
 Buy_{sing}-**se** always too-many sausages at the butcher shop Sanzot
"One (people) always buy too many sausages at the Sanzot butcher shop."

The (null) subject in (42) is understood as being indefinite, i.e. it receives a quasi-existential reading, whereas the subject in (43) is understood as a prototype, i.e. it receives a quasi-universal reading. Raposo & Uriagereka call the construction in (42) the **indefinite SE construction** (Belletti's passive *si* construction, roughly correspondent to Cinque's [+arg] *si*) and the construction in (43) is denominated the **generic SE construction** (Belletti's impersonal *si* construction, roughly correspondent to Cinque's [-arg] *si*).

Raposo & Uriagereka argue that the analyses of Otero's (1986) and Mendikoetxea's (1992) for the generic SE construction are right. However, they reject the previous analyses of the indefinite SE construction. Their work focuses on this issue and give an alternative explanation.

Raposo & Uriagereka argue that the agreeing DP in the indefinite SE construction is not in [Spec, T] (nor associated to this position). In other words, it is not a subject but rather it is hosted in a left-peripheral position (or associated to such position).

They argue that *se* in the indefinite SE construction is a syntactically active external argument (in a similar way than Cinque's [+arg] *si*), which is a morphologically defective determiner (it lacks person and number features for verbal agreement) with a Case feature to be checked. The Case that checks *se* in this construction is null Case²⁰ in

²⁰ In this sense, *se* is a type of minimal DP both semantically (reducing to the properties [+human] and [+indefinite]) and syntactically. It forms a natural class with PRO, which despite its being an argument, it

a [+finite] T(ense) lacking person and number Agr features. *Se* is generated in [Spec, vP] as a standard subject and moves to [Spec, TP] so as to check the D feature of T, and there it checks its null Case feature against the [+finite] T lacking person and number agreement features. Hence, the position [Spec, TP] is already occupied by *se* and then unable to host the agreeing DP.

Dobrovie-Sorin (1998, 2006) argues that certain pro-drop Romance languages (Italian, Spanish and European Portuguese, but not French and Romanian) have a subject *se/si* similar to the pronouns *one* in English, *on* in French, *man* in German and *men* in Dutch. This is nominative *se*.

This instance of *si* is different from ACC *si* (see section 4.4.1 below) in its Case features (it is marked for nominative rather than for accusative). Moreover, and unlike ACC *si*, NOM *si* is not an anaphor in the sense of Reinhart & Reuland (1993): it does not mark the predicate as reflexive and thus does not force coindexation between the subject and the object positions. NOM *si* is a subject clitic that binds an empty category in subject A-position, unlike ACC *si*, which does not bind a genuine trace.

Under the light shed by these data, Dobrovie-Sorin argues that in modern Romance languages, NOM *si* and ACC *si* appear to be two completely distinct linguistic entities, as different from each other as *sich* and *man* are in German, or *se* and *on* in French, and hence, no unified analysis is possible. NOM *si* and ACC *si* should be, therefore, treated as homonyms rather as instantiations of the same element. The availability of NOM *si* is regulated by a parameter that is positive for Italian, Spanish and European Portuguese, and negative for French and Romanian.

This is precisely the hypothesis put forward by Dobrovie-Sorin: Romanian lacks NOM *si* like French does. This is evidenced by the Romanian (44) counterparts of the Italian sentences, which are assumed to rely on NOM *si*.

- (44) a. *Nu **se** este niciodată mulțumit. (Dobrovie-Sorin 1998:405)
 Not **se** is never satisfied
 "One is never satisfied."
 b. *Adesea **se** este trădat de prieteni falși. (Dobrovie-Sorin 1998:405)
 Frequently **se** is betrayed by friends false
 "One is frequently betrayed by false friends."

is not specified for a full fledged Case but rather for Chomsky & Lasnik's (1993) Null Case (Raposo & Uriagereka 1996:780).

- c. *In această universitate **se** predă științele umane. (Dobrovie-Sorin 1998:405)
 In this university **se** teaches the humanities
"In this university one teaches the humanities."
- d. (Științele umane) le **se** predă în această universitate. (Dobrovie-Sorin 1998:405)
 (The humanities) them **se** teaches in this university
"(The humanities), one teaches them in this university."

There exists also a contrast between Spanish impersonal *se* and Romanian (45), which are thought to rely on NOM *si* too.

- (45) a. En esta escuela **se** castiga a los alumnos. (Dobrovie-Sorin 1998:405)
 In this school **se** punishes to the students
"In this school they punish the students."
- b. *In școala asta **se** pedepsește pe elevi. (Dobrovie-Sorin 1998:405)
 In school this **se** punishes the students
"In this school they punish the students."

4.2.2. Recapitulation and hypotheses

We have reviewed the most important analyses of impersonal and passive *se/si* in Romance, which can roughly be divided in two groups. The first is composed of those authors who think that the *se/si* clitic is an argument or forms a chain with an argument, and as such, is subject to the Theta Criterion (Belletti, Burzio, Manzini, Cinque for the cases of [+arg] *si*, Dobrovie-Sorin for impersonal *se/si*, Raposo & Uriagereka). On the other hand there are other authors for whom the clitic *se* is not an argument but marks something on the INFL head (Otero, Cinque for the cases of [-arg] *si*, Mendikoetxea, Dobrovie-Sorin for passive *se/si*).

The hypothesis that I will further develop in chapter 7 is that in both impersonal and passive *si* constructions, the clitic *si* is the spell-out of a defective anaphor in subject position (in fact, I will argue it is a SE-anaphor in the sense of Reinhart & Reuland 1993 and Reuland 2001²¹). Being this anaphor [-R], i.e. it cannot be referential by itself; it is semantically interpreted by means of a choice function that ranges from existential to universal readings. Arbitrary control will be accounted for on similar grounds.

²¹ Like Dobrovie-Sorin (1998, 2006) and Rivero (2001) do.

4.3. Middle SE

I will follow Ackema & Schoorlemmer (2006:132) to define the middle constructions that will be briefly dealt with in this section. For a middle construction holds the following statements:

- a) The external argument of the non-middle counterpart of the middle verb cannot be expressed as a regular DP-argument in the middle.
- b) If the non-middle counterpart of the middle verb has a direct internal argument role, the subject of the middle sentence carries this role.
- c) The middle verb is stative, non-episodic. The middle sentence is a generic statement. It expresses that the argument mentioned in (b) has a particular individual-level property, or that events denoted by the verb or the verb-argument combination have a particular property in general.

Some examples of middle sentences (Ackema & Schoorlemmer 2006):

(46) Bureaucrats bribe easily. *(English; Ackema & Schoorlemmer 2006:133)*

(47) La Tour Eiffel se voit de loin. *(French; Ackema & Schoorlemmer 2006:133)*
The Tower Eiffel *se* sees from afar
"The Eiffel Tower can be seen from afar."

(48) Dit boek leest als een trein. *(Dutch; Ackema & Schoorlemmer 2006:133)*
This book reads like a train
"This book is very easy to read."

4.3.1. Literature review

Manzini (1986) uses the term "middle *si*" to refer to the instance of *si* in Italian sentences such as (49). She states that this instance of *si* has a passivizer property that differentiates it from both the reflexive (see section 4.4) and the impersonal *si* (see section 4.2). In addition to this, sentences (50) and (51) shows an instance of what Manzini calls middle-reflexive *si*. However, these sentences that Manzini analyses as middles do not fit with the definition of middle given right above. (49), (50) and (51) do not talk about a property of the children or any other individual. Moreover, the verb "wash" is not a stative verb and the sentences do not have a generic reading.

(49) I bambini si lavano volentieri. (Manzini 1986:256)

The children si wash_{middle} gladly

"The children wash gladly."

(50) Gli unici bambini lavatisi. (Manzini 1986:258)

The only children (who) washed (themselves)

(51) Se ne lavano molti. (Manzini 1986:259)

Se of them wash (are washed/washed themselves) many

Nonetheless, Manzini give the definition in (52) as a unified definition with a new parameter (whether it has a passivizer function or not) that accounts for the so far four identified instances of *si*.

(52) Lexical entry of middle *si*: (Manzini 1986:257)

a. Variable: free.

b. Argument.

c. Categorial feature: N.

d. Φ -features: third-person, unspecified number and gender.

e. Clitic on a verb.

f. Bound to its subject.

g. Passivizer: yes.

Cinque (1988) argues that middle *si* like in (53) is a instance of this clitic in transitive contexts with verb-object agreement and it is [-arg] *si*. It is only possible with generic time reference and it requires some kind of manner adverb modification. It need not be associated with nominative, hence its compatibility with infinitival control structures. It is a pure [-arg] passivizer in the sense that it eliminates the external theta role (impersonal passive *si* retains it) to the extent that it is impossible to reassign it, as well as the accusative Case. Cinque calls this middle *si*. In other words, middle *si* is an [NP, IP] clitic and a "passivizer" [-arg] *si* that suspends the external θ -role and accusative Case.

(53) a. Neanche il nemico si uccide senza rimorsi. (Cinque 1988:560)

Not even the enemmy *si* kills without remorse

"Not even the enemy *si* kills without remorse."

- b. Carlo si odia facilmente. (Cinque 1988:560)
 Carlo *si* hates easily
 "Carlo *si* hates easily."
- c. Mario si festeggia con estrema difficoltà. (Cinque 1988:560)
 Mario *si* celebrates with extreme difficulty
 "Mario *si* celebrates with extreme difficulty."

Marelj (2004) studies middles in several languages (54), Romance among them. She follows the reasoning of Reinhart & Siloni (2005) in that the clitic *se/si* in Romance is a Case-absorbing morphology that absorbs accusative Case and it is not associated with a theta role. It is required when a lexical operation affecting the arity of verb takes place. The arb value of the middle is reached at LF, in line with the proposal of Chierchia (1995).

- (54) a. Porcelain vases break easily. (English; Marelj 2004:1)
 b. Porcelanska vaza **se** laki razbija. (Serbian/Croatian; Marelj 2004:2)
 Porcelain vases *se* break easily

Dobrovie-Sorin (1988, 2006) assumes that middle *se/si* is an instance of ACC-SE, which will be explained in more detail in section 4.4.1 below.

4.3.2. Recapitulation

To sum up, Manzini and Cinque think both that middle *si* is a clitic coindexed with subject position and whose function is a passivizer one, i.e. it eliminates the external θ -role (instead of retaining it like impersonal or passive *si*), as well as the accusative Case. On the other hand, Marelj argues that the clitic *se/si* is just a Case absorbing morphology and it is not associated with a θ -role.

4.4. Reflexive SE

Reflexive expression are those in which the subject and the object refer to the same entity. Examples of such verbs are "wash", "comb" and "shave"²² in (55):

²² Note however that these verbs can also have a transitive non-reflexive use:

- (i) Mary washed John
- (ii) Mary combed John
- (iii) Mary shaved John

- (55) a. John shaved himself.
 b. John combed himself.
 c. John combed himself.

4.4.1. Literature review

Burzio claims that reflexive *si* in (56) is a reflexive object clitic base-generated in clitic position, and it forms a chain with an empty category *ec* in object position, as other cases of object clitics. The verb assigns a theta role to *ec* as complement of the verb, and this θ -role is transmitted to the clitic via the chain. The clitic is the spell-out of the Case-marking features of the verb (Burzio 1986:38). The difference between *si* in (57a) and *lo* in (57b) is that *si* has an antecedent unlike *lo*, which has no antecedent in the sentence.

- (56) Maria **si** guarda. (Burzio 1986:37)
 Maria **si** watches
"Maria watches herself"

- (57) a. Maria **si** guarda *ec*. (Burzio 1986:37)
└──────────┘
"Maria watches herself"
- b. Maria **lo** guarda *ec*. (Burzio 1986:37)
└──────────┘
"Maria watches him"

Some verbs are obligatory reflexives, such as (58) and (59). Despite the presence of the clitic *si*, these verbs have objects and *si* cannot alternate with them (58a and 59a vs. 58b and 59b). Burzio assumes that these verbs are like idioms. Hence, they are highly idiosyncratic and probably codified in the lexicon with the requirement of *si*.

- (58) a. (Le vacanze) Giovanni **si** le sogna. (Burzio 1986:41)
 (The vacation) Giovanni **si** them dreams.
"(As for vacation) Giovanni dreams about it."
- b. *Giovanni gliele sogna. (Burzio 1986:41)
 Giovanni to-him-them dreams.
- (59) a. (La spiaggia) Giovanni **si** la immagina. (Burzio 1986:42)
 (The beach) Giovanni **si** it imagines.
"(As for the beach) Giovanni imagines it."

- b. *Giovanni gliela immagina. (Burzio 1986:42)
 Giovanni to-him-it imagines.

Manzini (1986) notes that reflexive *si* has a similar interpretation of controlled PRO (in its non-reciprocal reading), which leads Manzini to state that both elements introduce a dependent variable. Reflexive *si* is an anaphor subject to Chomsky's (1981) Condition A and furthermore, it has to be bound by a subject (never by an object). As impersonal *si*, reflexive *si* is considered an argument subject to the Theta Criterion with an N categorial feature and thus subject to the Case Filter. It also has unspecified number and gender features, and it is a clitic on a verb.

She argues that reflexive *si* can be summarized in (60).

- (60) Reflexive *si*: (Manzini 1986:251)
- a. Variable: dependent.
 - b. Argument.
 - c. Categorial feature: N.
 - d. Φ -features: third person, unspecified number and gender.
 - e. Clitic on a verb.
 - f. Bound to its subject.

Cinque (1998) argues that *si* in constructions such as (56) above is a true reflexive clitic that is an [NP,IP] clitic [+arg] that absorbs the external θ -role and the VP-internal Case.

Mendikoetxea (1997) claims that *se/si* clitics (at least in reflexive and arbitrary constructions) are the morphological realization of the [0]-person feature of AGR heads, which are "inert" for Case checking. The author defends an analysis in which structures with reflexive *se/si* involve the presence of PRO in the VP-internal subject position and have derived subjects, which accounts for the unaccusative-like properties of these constructions.

- (61) Los niños_i PRO se lavan t_i. (Mendikoetxea 1997:84)
 The children_i PRO se wash_{3p,pl} t_i.
 "The children wash."

Reflexive interpretation obtains when the internal argument moves to a position where it controls PRO, an operation forced by the fact that AgrO is "inert" for ACC Case-checking. That the internal argument must be obligatorily externalized explains why reflexive constructions are incompatible with processes that require that the internal

argument remain within the VP, such as *ne*-cliticization in Italian and bare plurals. This control analysis accounts as well for why reflexive structures must be (at least) bi-argumental: we need at least two arguments, the *controller* and the *controlee*. The element *se/si* is not an anaphor (there is not such a thing as *reflexive se/si*) but it is an agreement morpheme which has a [0-person] feature (like Mendikoetxea's (1992) analysis of impersonal *se*).

According to Reinhart & Sioni (2005), reflexive verbs as in (56) are the result of a reflexivization operation at the syntax (in syntax-languages as the majority of Romance languages). This operation bundles the agent and the theme θ -role in the external element. The verb behaves as an unergative verb (Reinhart & Sioni 1999, see also Dobrovie-Sorin 2006). The clitic *si* is just a morphological marker that absorbs the accusative Case and informs that the operation of reflexivization has taken place.

Dobrovie-Sorin (1998, 2006) argues that all Romance languages have an anaphoric reflexive-reciprocal (object) *se/si* as shown in (62) for French. She calls this instance of *se* **accusative se** (ACC *se* hereinafter).

- (62) a. Jean **se** lave. (Dobrovie-Sorin 2006:123)
 Jean **se** washes
"Jean washes himself."
- b. Le grec **se** traduit facilement. (Dobrovie-Sorin 2006:122)
 the Greek **se** translates easily
"Greek translates easily."
- c. Il s'est traduit trois romans. (Dobrovie-Sorin 2006:122)
 It **se**-has translated three novels
"Three novels were translated."
- d. La branche s'est cassée. (Dobrovie-Sorin 2006:121)
 The branch **se**-has broken
"The branch broke."
- e. Jean s'est souvenu de Marie. (Dobrovie-Sorin 2006:169; endnote 10)
 Jean **se**-has remembered of Marie
"Jean remembered Marie."

This instance of *se* is witnessed in reflexive, middle, passive (with *se*), ergative and inherent verbs (see (62)). The common characteristics of *se* in all the examples of (62) are that, first *se* is an object clitic marked with accusative Case, and second, *se* is an

anaphor in the sense of Reinhart & Reuland's (1993)²³, in other words, *se* is a reflexive marker that forces the coindexation of the subject and object positions. ACC *se* does not bind a genuine trace, and it is based-generated in an A'-position, in line with the claim that it is a reflexive marker in the sense of Reinhart & Reuland (1993). Note that only Dobrovie-Sorin and Rivero (2001)²⁴ consider (ACC) *si* as an anaphor. They claim that ACC *si* has an anaphoric status, as it is understood in the framework of Reinhart & Reuland (1993). So will I claim but not only for ACC *si* but for all instances of *se/si*.

4.4.2. Recapitulation and hypotheses

Burzio, Manzini and Cinque argue that reflexive *si* is an argument (or a clitic forming a chain with an argument) theta and case-marked. In this sense, the reflexive verbs with the clitic are transitive verbs. On the other hand, Dobrovie-Sorin conceives reflexive *se* as an anaphor in the sense of Reinhart & Reuland, and it marks the predicate as reflexive, forcing the coindexation of subject and object positions. Reinhart & Siloni state that the clitic is just a morphological marker that absorbs the accusative Case and marks the application of the syntactic operation of reflexivization by bundling two θ -roles in the external argument. Mendikoetxea argues that reflexive *se/si* is the morphological realization of the [0]-person feature of the AgrO head, which renders it "inert" for Case checking. Hence, PRO can be in VP-internal subject position and the internal arguments moves up to AgrS, whence it controls the clause internal PRO.

The hypothesis that I will develop in chapter 5 is that reflexive *se* is a SE-anaphor that is inserted not to realize a θ -role but due to formal conditions so that the syntactic derivation converge at the interface with the C-I system.

4.5. Inchoative / ergative SE

With the term *inchoative / ergative verb* I refer to "*a special kind of intransitive verb. Semantically, its subject does not actively initiate or is not actively responsible for the action of the verb; rather, it has properties which it shares with the direct object of a transitive verb (or better, with the grammatical subject of its passive counterpart). (Kersten, Ruys & Zwaarts 2001)*". These verbs participate in the causative-inchoative

²³ In chapter 4 I will describe Reinhart & Reuland's (1993) theoretical framework in more detail, as well as SE-anaphors and SELF-anaphors.

²⁴ See section 4.7.2.

alternation, as can be seen for the verbs "open", "break" and "melt" in (63), (64) and (65) respectively.

- (63) a. The door opened. *(inchoative)*
 b. John opened the door. *(causative)*
- (64) a. The glass broke. *(inchoative)*
 b. John broke the glass. *(causative)*
- (65) a. The ice melted. *(inchoative)*
 b. The son melted the ice. *(causative)*

4.5.1. Literature review

Burzio (1986) argues that ergative *si* in (66) is a morphological reflex of the "loss" of subject θ -role, which marks the ergative derivation in order to distinguish it from its transitive counterpart. He claims that this is an idiosyncratic lexical process. Ergative *si* plays no syntactic role at all.

- (66) Il vetro **si** rompe. *(Burzio 1986:37)*
 The glass **si** breaks
 "*The glass breaks (itself)*"

According to Manzini (1986), the ergative *si* identified by Burzio (1986) as that in sentences (66) is the same *si* as that one defined in (52) for middle *si*. The difference is that the clitic attaches to the verb at the lexicon rather than at the syntax.

- (67) Lexical entry of middle *si*: *(Manzini 1986:257)*
- a. Variable: free.
 - b. Argument.
 - c. Categorical feature: N.
 - d. Φ -features: third-person, unspecified number and gender.
 - e. Clitic on a verb.
 - f. Bound to its subject.
 - g. Passivizer: yes.

Cinque (1988) states that ergative *si* is a reflexive [NP,IP] clitic that is [-arg] and suspends the external θ -role and VP-internal Case.

Burzio (1986) argues that ergative *se* like in Spanish examples (68), is a morphological marker that has no function in the syntax. He further argues that ergative

se undergoes affixation in the lexicon and hence its syntactic inactivity (contrary to impersonal *si* that undergoes syntactic affixation). Furthermore, ergative *se* does not involve the interpretation of a non-specified (indefinite) human subject unlike impersonal *si* (Otero 1986).

(68) a. La puerta **se** abrió (por sí sola).

The door **se** opened (by itself)

"The door opened (by itself)."

b. El hielo **se** fundió.

The ice **se** melted

"The ice melted."

Masullo (1999) focuses on Burzio's *si* and studies the unaccusative verbs in Spanish. More concretely, he addresses the issue that some of these verbs require the clitic *se*, whereas others prevent its occurrence, and moreover, some others allow it though do not require it. In other words, he addresses the issue of the ergative *se* in Spanish.

Masullo makes a first distinction between *derived unaccusative verbs* like *romperse* (break), which require the clitic *se* and have a transitive counterpart, and *inherent unaccusative verbs* like *llegar* (arrive), which do not allow the presence of the clitic *se* nor have a transitive alternative, and they are usually existential and presentational verbs.

Masullo follows Levin & Rappaport (1995) in that the causative derivations are basic (listed in the lexicon), whereas the ergative constructions are formed from the causatives. However, he differs in saying that the formation of ergative derivation takes place in the syntax rather than, as Levin & Rappaport claim, in the lexicon. He develops an analysis based on incorporation. The ergativization takes place in the syntax by means of the insertion of the clitic *se*, which is a nominal head that incorporates in an aspectual head above the verb and absorbs accusative. Hence, the remaining nominal (the theme) needs to check nominative Case against Tense in order to fulfil the Visibility Condition and the Case Filter. The clitic *se* denotes an external argument with very little semantic specification, compatible either with internal causation or with inanimate external causation. Hence, it is incompatible with animate external causation.

Finally, Reinhart & Siloni (2005) claim that the role of the clitic *si/se* with ergative verbs is similar to its function with reflexive verbs. Ergative verbs are derived verbs by a lexical operation (decausativization) that reduces the arity of verb by

reducing the external θ -role. The clitic *se/si* is just a morphological marker of the fact that a reduction operation has taken place in the lexicon.

4.5.2. Recapitulation and hypotheses

Burzio claims that ergative *se/si* is just a morphological reflex of the loss of the subject θ -role. Also Reinhart & Siloni claim that the clitic is just a morphological mark that informs about a decausativization lexical operation that has rendered the causative entry in a unaccusative one. Manzini, however, assumes that ergative *si* is an argument N that is cliticized, but it has a passivizer function. Cinque claims that it is a clitic but [-arg], this means that it suspends (does not bear) the external θ -role and VP-internal Case. Masullo assumes that it is a clitic NP that incorporates in an aspectual head, absorbs accusative and denotes an external argument semantically specified.

The hypothesis that I will develop in chapter 6 is that ergative *se* is a SE-anaphor (like reflexive *se*) that is inserted not to realize a θ -role but rather due to formal conditions so that the syntactic derivation converge at the interface with the C-I system.

4.6. Inherent SE

With the term of *inherent se* I refer to the instance of *se* that appears with the so-called *inherent reflexive verbs*, which have the same syntactic and semantic properties as the ergative verb with the notable exception that they do not participate in the causative-inchoative alternation. These are verbs such as the Spanish verbs *arrepentirse* ("change of mind") and *desmayarse* ("faint").

4.6.1. Literature review

Burzio (1986) claims that inherent-reflexive *si* in (69) appears with verbs that do not participate in the causative alternation. There is no reflexive object and these verbs are unaccusative, as *ne*-cliticization indicates. Burzio argues that this instance of *si* is, like ergative *si*, just a morphological marker. The difference between both is that the former does not have a causative alternation.

- (69) Giovanni **si** sbaglia. (Burzio 1986:37)
 Giovanni **si** mistakes
 "Giovanni mistakes (himself)"

Cinque (1988) claims that inherent *si* is an [NP,IP] clitic that is [-arg] and simply marks the absence of external θ -role and VP-internal Case.

Reinhart & Siloni (2005) claim that the role of the clitic *si/se* in sentences such as (69) above is similar to its function with ergative verbs. These verbs are ergative verbs derived verbs by a lexical operation (decausativization) that reduces the arity of verb by reducing the external θ -role. The clitic *se/si* is just a morphological marker of the fact that a reduction operation has taken place in the lexicon. The peculiarity of these verbs is that their causative alternation is not grammaticalized.

Masullo (1999) makes a first distinction between *derived unaccusative verbs* like *romperse* (break), which require the clitic *se* and have a transitive counterpart, and *inherent unaccusative verbs* like *llegar* (arrive), which do not allow the presence of the clitic *se* nor have a transitive alternative, and they are usually existential and presentational verbs. Within the group of derived unaccusative verbs, there is a subset of verbs that do not have a transitive counterpart. Masullo calls these **inherent ergative verbs**, and they generally express changes of position, disposition or physical or mental state, like *arrepentirse* (change one's mind) and *enrojarse* (redden). Since these verbs can form adjectives, Masullo argues that they have two verbal layers (one expressing the cause and the other the change of state) although they are compatible only with internal causation (unlike the derived unaccusative verbs). This can be seen with *con-expansion*, which cannot introduce an external force. The only way to express this causation is by means of *se*, which is coindexed with the internal argument. In a wide sense, this is a reflexive construction because both arguments are coindexed. However, only reflexive *se* admits expansion by means of insertion of a *a sí mismo* (himself) phrase.

Rigau (1994) studies pronominal verbs in Catalan like *desmaiarse* (faint), which have an incorporated anaphoric pronoun. She focuses on the aspectual characteristics of these verbs. She notes the perfective value of these verbs and her proposal consists of three claims. First, pronominal verbs always have an internal argument. Second, the incorporated clitic triggers determined syntactic and semantic restrictions. And third, the perfective aspect of the predicates denoted by the pronominal verbs is a consequence of the benefactive character of the clitic.

Rigau claims that the clitic *se* with accusative verbs absorbs the inherent partitive case, which is responsible for the non-specific reading of the object and for the licensing of bare NPs. Hence, the theme has to be licensed by accusative assignment, which triggers a specific reading.

When the verb does not assign partitive nor accusative but dative, the clitic absorbs it as with the verbs in (70). When the verb does not assign an inherent case, the clitic absorbs the accusative Case as in the case of *lamentarse* in Spanish, and the object needs a preposition like *de* (of) in (71) which is an extra case assigner.

(70) a. El teu menyspreu dol a la Maria. (Rigau 1994:34)

The your underestimation hurts to the Maria
"Your underestimation hurts Maria."

b. La Maria es dol de teu menyspreu. (Rigau 1994:34)

The Maria **es** hurts of your underestimation
"Your underestimation hurts Maria."

c. En Pere ha agradat a la Maria. (Rigau 1994:34)

The Pere has pleased to the Maria
"María has liked Pere."

d. La Maria s'ha agradat d'en Pere. (Rigau 1994:34)

The Maria **se**-has pleased of-the Pere
"María has liked Pere."

(71) a. María lamenta eso.

María regrets that
"María regrets that."

b. María **se** lamenta de eso.

María **se** regrets of that
"María regrets that."

With unaccusative verbs, which are thought to assign partitive but not accusative, the clitic absorbs the partitive case and hence, only nominative Case is available for the theme and bare plurals are banned as in (72).

(72) a. Han florit (els) cirerers. (Rigau 1994:34)

Have blossomed (the) cherry trees
"The cherry trees have blossomed"

b. S'han florit *(els) formatges. (Rigau 1994:34)

Se-have deteriorated *(the) cheeses
"The cheeses have deteriorated"

The claim is that the clitic has a nominal character and hence, it requires a θ -role. It usually realizes the benefactive θ -role, which triggers the perfective interpretation. Finally, she claims that the clitic is anaphoric, i.e. without an independent referential value. Therefore, it raises to AgrS and acquires the ϕ -features (person, number and gender) of the argument that occupies the subject position.

De Miguel & Fernández Lagunilla (2000) show that the clitic *se* that appears with certain unaccusative and transitive verbs in Spanish, such as in the examples of (73), is an operator that focuses a phase of the event denoted by the predicate in which it appears²⁵. In that sense, the clitic is quite similar to some aspectual adverbs like *aún* (yet), *todavía* (still/yet) and *ya* (already), which are thought to be aspectual operators too.

- (73) a. Juan **se** murió ayer. (De Miguel & Fernández Lagunilla 2000:13)²⁶
 Juan **se** died yesterday
"Juan died yesterday."
- b. El libro **se** cayó del estante. (De Miguel & Fernández Lagunilla 2000:13)
 The book **se** fell from the shelf
"The book fell off from the shelf."
- c. Juan **se** bebió una caña/*cerveza. (De Miguel & Fernández Lagunilla 2000:13)
 Juan **se** drank one beer/*beer
- d. Juan **se** ha visto toda la película/*cine inglés.
(De Miguel & Fernández Lagunilla 2000:13)
 Juan **se** has seen all the film/*english cinema
"Juan has seen the film (completely)/english cinema."

They reject previous hypotheses such as Rigau's (1994), which claimed that *se* with these kinds of verbs is a lexical element that marks *perfectivity*, on the ground of examples in (74), which clearly show that something more than perfectivity is needed in order to license the clitic. Their alternative hypothesis is that the distinction of *perfectivity* and *culmination* can account for the grammaticality of (73) vs. the agrammaticality of (74).

²⁵ The term "phase" in the work of De Miguel & Fernández Lagunilla (2000) is not to be confused with the use that Chomsky (2001) and subsequent works make of the same term. In this section, the term makes reference to a part of the event (a sub-event).

²⁶ Glosses and translations of all the examples of De Miguel & Fernández Lagunilla's (2000) are mine.

(74) a. *El niño **se** nació sietemesino.

(De Miguel & Fernández Lagunilla 2000:13)

The child **se** was-borned two months premature

"The child was borned two months premature."

b. ***Me** vi la costa y **me** dirigí a ella.

(De Miguel & Fernández Lagunilla 2000:13)

Me_{1st,sing} saw the coast and **me**_{1st,sing} went to it

"I saw the coast and I went to it."

c. *El agua **se** hirvió en un instante. (De Miguel & Fernández Lagunilla 2000:13)

The water **se** boiled in a while

"The water boiled in a while."

d. *María **se** engordó (dos kilos). (De Miguel & Fernández Lagunilla 2000:13)

María **se** gained-weith (two kg.)

"María gained weith (gained two kg)."

Their analysis uses Pustejovsky's (1991) typology of events, which defines three types of events (states, processes and transitions). The processes and transitions are complex events in the sense that they have phases (parts), in which the event may initiate, culminate or end.

The hypothesis that these linguists put forward is that the clitic *se* is an aspectual operator that signals that the event culminates in a given point, in which a change of state takes place.

To sum up, De Miguel & Fernández Lagunilla defends that *se* is an aspectual operator, which requires a complex eventive structure that denotes an achievement that culminates in a point in which a change of state takes place. Therefore, they call this clitic *culminative se*.

4.6.2. Recapitulation and hypotheses

Burzio states that inherent *se* is a morphological reflex of the loss of the subject θ -role, like ergative *si*. The difference is that verbs marked with inherent *se* do not have a causative alternation. Cinque follows the same line and assumes that *si* is a[NP,IP] [-arg] clitic that marks the absence of external θ -role and VP-internal Case. Also Reinhart & Siloni assume that *se/si* is a morphological marker of a decausativization operation in the lexicon, although the difference with unaccusative verbs that enter in the inchoative-causative alternation, is that these verbs have not lexicalized their causative entry.

Masullo, instead, assumes that *se* is an argument that is coindexed with the subject position so as to express internal causation, in this sense, it is a reflexive. Rigau (1994) considers *se* an incorporated anaphoric pronoun that requires θ -role and absorbs Case (accusative) or case (partitive). Finally, De Miguel & Fernández Lagunilla (2001) claim that *se* is an aspectual operator whose function is to denote an achievement that culminates in a point in which a change of state takes place.

The hypothesis that I will develop in chapter 6 is that inherent *se* is a SE-anaphor (like reflexive and ergative *se*) that is inserted not to realize a θ -role but rather due to formal conditions so that the syntactic derivation converge at the interface with the C-I system.

4.7. Recapitulation and implications

4.7.1. Recapitulation

In tables (75)-(79) below, a recapitulation of the different works on the status and function of the clitic *se/si* in Romance can be found. Table (75) summarizes the so far reviewed works on impersonal *se/si*, while table (76) does the same with the works on passive *se/si*. Table (77) summarizes works on middle *se/si*, (78) on reflexive *se/si*, (79) lists the works on inherent *se/si* and finally table (80) concludes with the works on ergative *se/si*.

(75) Impersonal *se/si*:

Author	Language	Status and function of <i>se/si</i>	
Belletti (1982)	Italian	status	Clitic pronoun (pronominal) based-generated in INF.
		function	It absorbs NOM and bears the external θ -role, so that it renders INFL a proper governor able to govern an empty category in [Spec,IP].
Burzio (1986)	Italian	status	Subject clitic based-generated in any NP position that cliticizes from [Spec,IP] position and forms a chain with an empty category in [Spec,IP].
		function	The chain [ec, <i>si</i>] bears nominative Case and the external θ -role.
Manzini (1986)	Italian	status	Argument Clitic on the verb with third person and unspecified number and gender ϕ -features.
		function	It is bound to subject position, and it introduces a free variable.
Otero (1986)	Spanish	status	Absorber clitic, which is a marker of INFL.
		function	It absorbs the definiteness of a finite INFL so that pro_{arb} can be introduced in subject position.
Cinque (1988)	Italian	status	Clitic pronoun coindexed with the subject position and with Agr. It can be an argument or not (in which case is just a morphological marker)
		function	[+arg] <i>si</i> is associated to θ -role and NOM, [-arg] need not be associated with theta nor Case and its function is to provide personal Agr with the features able to identify the content of pro as an unspecified generic person pronominal.
Mendikoetxea (1992)	Spanish	status	It is the spell out of the feature person of the AgrS projection.
		function	it absorbs NOM Case but not θ -role in impersonal, so that PRO can be inserted in subject position and realize the external θ -role
Mendikoetxea (2008)	Spanish	status	AGR-clitic specified for 0-person
		function	It lacks a D-feature that prevents it from checking the EPP-feature of T and checks the person and number feature of T and turns it into a non-referential T, so that GnPro can be inserted.
Dobrovie-Sorin (1998)	French & Romanian	status	Anaphoric argument subject clitic.
		function	It binds an empty category in subject A-position.
Rivero (2001) ²⁷	Polish	status	SE-anaphor marked with NOM Case.
		function	Head a [+R] chain with a defective pronoun so that it can be interpreted as an indefinite following the analysis of Chierchia (1995).

²⁷ See section 4.7.2 below.

(76) Passive *se/si*:

Author	Language	Status and function of <i>se/si</i>	
Belletti (1982)	Italian	status	Clitic pronoun (pronominal) based-generated in INF.
		function	It absorbs ACC and bears external theta (similar to passive morphology).
Burzio (1986)	Italian	status	Subject clitic that is based generated in object position and thence, it moves to [Spec,IP] whence it cliticizes and forms a chain with an empty category in [Spec,IP].
		function	the chain [<i>ec,si</i>] bears accusative Case and the external θ -role
Mendikoetxea (1992)	Spanish	status	It is the spell out of the feature person of the AgrO projection.
		function	it absorbs ACC Case and the external θ -role.
Mendikoetxea (2008)	Spanish	status	AGR-clitic specified for 0-person
		function	it lacks a D-feature that prevents it from checking the EPP-feature of T and checks the person feature of T and turns it into a non-referential T, so that GenPro can be inserted. T checks its number person against the agreeing DP.
Raposo & Uriagereka (1996)	Portuguese	status	A morphologically defective determiner (it lacks person and number features for verbal agreement) with a Case feature.
		function	It is the external argument, it checks null Case against a [+fin] T lacking person and number Agr features. <i>Se</i> checks the D feature of T too, hence the agreeing subject is in left peripheral-position.

(77) Middle *se/si*:

Author	Language	Status and function of <i>se/si</i>	
Manzini (1983)	Italian	status	Argument Clitic on the verb with third person and unspecified number and gender ϕ -features.
		function	It is bound to its subject, it introduces a free variable, and it has a passivizer function.
Cinque (1988)	Italian	status	Clitic pronoun coindexed with the subject position and with Agr, it is [-agr].
		function	it is a pure [-arg] passivizer that eliminates the external θ -role (instead of retaining it, like impersonal or passive <i>si</i>) and accusative Case.
Marelj (2004)	Romance	status	Case absorbing morphology.
		function	It absorbs accusative Case and it is not associated with a θ -role. The arb value is reached at LF
Dobrovie-Sorin (1998)	Romance	status	ACC <i>si</i> : object clitic that is an anaphor (R&R 1993).
		function	It marks the predicate reflexive: i.e. it forces coindexation subject-object position: it is generated in A'-position and does not bind a genuine trace.

(78) Reflexive *se/si*:

Author	Language	Status and function of <i>se/si</i>	
Burzio (1986)	Italian	status	Reflexive (argument) object clitic based-generated in clitic position.
		function	It forms a chain with an empty category in object position and relates it with the subject, which is the antecedent of the clitic.
Manzini (1986)	Italian	status	Argument Clitic on the verb with third person and unspecified number and gender ϕ -features.
		function	It is bound to its subject, and introduces a dependent variable bound by the subject.
Cinque (1988)	Italian	status	True reflexive clitic [NP,IP] [+arg].
		function	It absorbs the external θ -role and the VP-internal Case.
Mendikoetxea (1997)	Spanish	status	Morphological realization of a [0]-person feature on the AgrO head.
		function	It renders AgrO "inert" for Case checking, hence the internal argument moves up to AgrS whence it control the clause-internal PRO in VP-internal subject position.
Reinhart & Siloni (2005)	Romance	status	Morphological marker.
		function	It absorbs the accusative Case and informs that the operation of reflexivization has taken place in the syntax.
Dobrovie-Sorin (1998)	Romance	status	ACC <i>si</i> : object clitic that is an anaphor (R&R 1993)
		function	it marks the predicate reflexive: i.e. it forces coindexation subject-object position: it is generated in A'-position and does not bind a genuine trace.

(79) Inherent *se/si*:

Burzio (1986)	Italian	status	Morphological reflex of the loss the subject θ -role.
		function	It marks the derivation as ergative though it has no causative alternation.
Cinque (1988)	Italian	status	Clitic [NP,IP] [-arg].
		function	It marks the absence of external θ -role and VP-internal Case.
Reinhart & Siloni (2005)	Romance	status	Morphological marker.
		function	It marks a verb as the result of a reduction operation at the lexicon, although the causative base entry has not been lexicalized in that language.
Masullo (1999)	Spanish	status	Argument clitic.
		function	<i>Se</i> is coindexed with the internal argument and in a wide sense, this is a reflexive construction because both arguments are coindexed (arguments of two verbal layers, one expressing the cause and the other the change of state; and they are only compatible with internal cause)

Rigau (1994)	Catalan Spanish	status	Incorporated anaphoric pronoun. Requires θ -role. Usually benefactive character
		function	Raises to AgrS, acquire the ϕ -features of the argument in subject position and absorbs inherent partitive case or accusative Case. the benefactive θ -role triggers the perfective interpretation
De Miguel & Fernández Lagunilla (2000)	Spanish	status	Aspectual operator
		function	It denotes an achievement that culminates in a point in which a change of state takes place (culminative <i>se</i>).

(80) Ergative or inchoative *se/si*:

Author	Language	Status and function of <i>se/si</i>	
Burzio (1986)	Italian	status	Morphological reflex of the loss the subject θ -role.
		function	It marks the ergative derivation so as to distinguish it from its transitive counterpart. <i>Si</i> plays no syntactic role at all.
Manzini (1986)	Italian	status	Argument Clitic on the verb with third person and unspecified number and gender ϕ -features.
		function	It introduces a free variable but at the lexicon instead of the syntax as middle <i>si</i> does.
Cinque (1988)	Italian	status	Reflexive [NP,IP] clitic [-arg].
		function	It suspends the external θ -role and VP-internal Case.
Mendikoetxea (1992)	Spanish	status	Morphological reflex of the loss the subject θ -role.
		function	It marks the ergative derivation so as to distinguish it from its transitive counterpart: it plays no syntactic role at all.
Masullo (1999)	Spanish	status	Argument clitic.
		function	It incorporates in an aspectual head and absorbs accusative, so that the remaining nominal checks NOM against INFL. <i>Se</i> denotes an external argument with very little semantic specification, compatible either with internal causation or with inanimate external causation.
Reinhart & Siloni (2005)	Romance	status	Morphological marker.
		function	It marks a verb as the result of an reduction operation at the lexicon.

4.7.2. ***Se* in Slavic**

Rivero (2001) notes that reflexive clitics also appear in Slavic languages in the uses that Dobrovie-Sorin calls ACC *se/si*, as can be seen in the Spanish and Polish examples in

(81): both languages have a reflexive use of *se/si/się* (81a,a'), a middle use (81b,b'), an ergative use (81c,c'), and an inherent or intrinsic use of *se/si/się* in (81d,d').

- (81) a. Juan **se** viste. (Rivero 2001:170)
 Juan **se** dresses
- a'. Janek ubiera **się**. (Rivero 2001:170)
 John dresses **się**
"John gets dressed."
- b. Este coche **se** conduce fácilmente. (Rivero 2001:170)
 This car **se** drives easily
- b'. Ten samochód prowadzi **się** łatwo. (Rivero 2001:170)
 This car drives **się** easily
"This car drives easily."
- c. El vaso **se** rompió. (Rivero 2001:170)
 The glass **se** broke
- c'. Szklanka **się** rozbiła. (Rivero 2001:170)
 Glass **się** broke
"The glass broke."
- d. María **se** asusta de Juan. (Rivero 2001:170)
 María **se** fears of John
- d'. Maria boi **się** Janka. (Rivero 2001:170)
 Maria fears **się** John
"Mary fears/is afraid of John."

The uses in (81) are attested both in Romance languages (French, Italian, Portuguese, Spanish and Romanian), and in Slavic languages (Polish, Bulgarian, Croatian/Serbian, Slovenian, Czech, Macedonian and Slovak). Except French, all these languages display the use of the reflexive clitic with intransitive verbs, like Polish in (82b).

- (82) a. Aquí **se** trabaja mucho. (Rivero 2001:170)
 Here **se** works much
- b. Tutaj **się** pracuje sporo. (Rivero 2001:170)
 Here **się** works much
"Here people works a lot."

What Rivero calls *nominative impersonal* is instantiated in (83a) for Spanish and in (83b) for Polish.

- (83) a. Antes **se** leía estos libros con placer. (Rivero 2001:171)
 Before **se** read_{3rd,sing} these books with pleasure
"In the past one/people read these books with pleasure."
- b. Tę książkę (czyta / czyta**o**) **się** z przyjemnością. (Rivero 2001:171)
 This book_{acc} (read_{3rd,sing} / read_{neu}) **się** with pleasure
"One (reads / read) this book with pleasure."

Some Romance and Slavic languages have the nominative impersonal, some other do not, as can be seen in table (84).

- (84) The nominative impersonal: (Rivero 2001:171)

	Yes	No
Romance	Italian, Portuguese, Spanish	French, Rumanian
Slavic	Polish, Slovenian	Bulgarian, Czech, Slovak, Serbo-Croatian (?)

Rivero proposes that a sentence with a nominative impersonal, such as (83), has a structure like (85).

- (85) [CIP [C_I *se/si/się*] [TP [T pres/past] [VP NP₁ V NP₂]]] (Rivero 2001:182)

The reflexive clitic *se/si/się* is a SE-anaphor (in the sense of Reinhart & Reuland (1993)) marked with nominative Case but without the reflexivizing function of the SELF-anaphors and [-R] (until NP₁ raises to [Spec,CIP]). The T(ense) head is defective in that it lacks ϕ -features. NP₁ is merged in subject position. It is a null defective pronoun that lacks ϕ -features and hence, the D-layer (i.e. it is an NP). This pronoun is minimally specified for the feature [+human] and has a Case feature to be checked. However, it cannot be checked against T since it is ϕ -defective, and thus it must raise to [Spec,CIP] to check nominative Case against the clitic *se/si/się*. By virtue of this movement, the clitic *se/si/się* heads a chain with the feature [+R] and then the defective pronoun is interpreted as an indefinite following the analysis of Chierchia (1995): it is a variable existentially bound, which can be disclosed by adverbials. The verb checks its ϕ -features against defective T and hence, the verb shows up with default morphology (third person singular in Spanish, third person singular or neuter in Polish).

4.7.3. Romance *se/si* and Germanic languages

Germanic languages, in particular Dutch, use SE-anaphors (Everaert 1986; Reinhart & Reuland 1993) in some constructions where Romance uses a *se/si* clitic, such as some reflexives (86a), reciprocals (86b) and inherent reflexives (87) but not inchoatives that have causative alternations (88):

(86) a. Jan wast **zich**.

Jan washes **zich**

"Jan washes."

a. Jan en Marie wassen **zich**.

Jan and Marie wash **zich**

"Jan and Marie wash each other."

(87) Jan heeft **zich** bedacht.

Jas has **zich** changed of mind

"Jan has changed his mind."

(88) De vaas is (***zich**) gebroken.

The glass is (***zich**) broken

"The glass has broken."

The hypothesis that I put forward in the following chapters is that the Germanic *zich* SE-anaphor is the same element in the lexicon as the Romance *se/si* clitic. Differences in the lexicon-syntax, syntax-phonology and syntax-semantics will account for the variation across languages despite the fact that this element is the same in all those languages (and ideally, in all languages, i.e. it is universal). Even in the case of languages like English, which lacks both overt SE-anaphors like *zich* and clitics like *se/si*, make use of this kind of anaphors though they are not phonologically realized.

4.7.4. From null to overt SE-anaphors

So far, we have seen control constructions in chapter 3 where a null SE-anaphor PRO is in subject position. In this chapter we have reviewed other constructions in Romance where the clitic *se/si* appears: reflexives and reciprocals, impersonals and passives, middles, inchoatives or ergatives and inherent ergatives. We have seen that also Slavic languages have a counterpart of the Romance *se/si* clitic and that Germanic languages make use of a SE-anaphor in some of the constructions reviewed in this chapter.

(89) SE-anaphors across languages and constructions:

	English	Dutch	Romance	Slavic
OC	null SE-anaphor (PRO)	null SE-anaphor (PRO)	null SE-anaphor (PRO)	null SE-anaphor (PRO)
NOC	null SE-anaphor (PRO)	null SE-anaphor (PRO)	null SE-anaphor (PRO)	null SE-anaphor (PRO)
AC	null SE-anaphor (PRO)	null SE-anaphor (PRO)	null SE-anaphor (PRO)	null SE-anaphor (PRO)
Impersonal	does not exist	does not exist	overt SE-anaphor (<i>se/si</i>)	overt SE-anaphor (<i>się</i>)
Passive	does not exist	does not exist	overt SE-anaphor (<i>se/si</i>)	overt SE-anaphor (<i>się</i>)
Middle ²⁸	-	-	-	-
Reflexive	null SE-anaphor (PRO' ²⁹)	overt SE-anaphor (<i>zich</i>)	overt SE-anaphor (<i>se/si</i>)	overt SE-anaphor (<i>się</i>)
Reciprocal	null SE-anaphor (PRO')	overt SE-anaphor (<i>zich</i>)	overt SE-anaphor (<i>se/si</i>)	overt SE-anaphor (<i>się</i>)
Ergative	null SE-anaphor (PRO')	null SE-anaphor (PRO')	overt SE-anaphor (<i>se/si</i>)	overt SE-anaphor (<i>się</i>)
Inherent reflexive	null SE-anaphor (PRO')	overt SE-anaphor (<i>zich</i>)	overt SE-anaphor (<i>se/si</i>)	overt SE-anaphor (<i>się</i>)

The hypothesis that will be defended in this dissertation is that all the so far seen constructions have in common the presence of a SE-anaphor, be this null, clitic or tonic, in Romance, Germanic and Slavic languages. I have already provided an account of control phenomena based on null SE-anaphors (PRO). In the following chapters, I will provide an account for reflexive, inchoative and inherent ergative constructions based on both null (PRO')³⁰ and overt SE-anaphors, as summarized in table (89).

²⁸ Note that I will not deal with middle *si* and middle constructions in this dissertation, though. I will leave this instance of *si* as a question open to future research.

²⁹ I will introduce the null SE-anaphor PRO' in contrast with PRO in chapter 4.

³⁰ See footnote 29.

Chapter 5

Reflexivity and overt SE-anaphors at the interfaces

I will show in this chapter that Spanish makes use of both complex (SELF-) and simple (SE-) anaphors. This is important because I will demonstrate that the clitic *se* in Spanish is the morphological realization of a null SE-anaphor, and in this sense, I will argue that both Romance and Germanic languages form reflexives in a similar way.

As I said in chapter 3 (section 3.3.2.3.), SE-anaphors do not obey either Condition A or Condition B proposed by Chomsky (1981). Therefore, they can be either locally bound, as I will argue in this chapter that it happens with *inherent reflexive verbs*, or non-locally bound (though this possibility is not attested in Spanish, i.e. binding in Reuland & Koster's (1991) domains 2 and 3). SELF-anaphors in Spanish (which are formed following the pattern *x+mismo*) are necessary to license the reflexive reading of *non-inherent reflexive verbs*. The clitics that appear with inherent reflexive verbs are SE-anaphors inserted along the syntactic derivation as a last resort mechanism in order for the derivation to converge at the C-I interface. The clitics are needed to adjust the valence (arity) of the verb and the formal requirements of the syntax. In conclusion, I will show that Reinhart & Reuland's (1993) A and B Conditions hold for English, Dutch and Spanish. The cross-linguistic variation in the occurrence of SE-anaphors with inherent reflexive verbs will be explained in terms of conditions on the spell-out of the ϕ -features of SE-anaphors (i.e. by resorting to mechanisms at the S-M interface).

The chapter is structured as follows. The first section is devoted to presenting the empirical data regarding reflexivization in English, Dutch and Spanish. Subsequently, the theoretical background will be exposed. In the third section I will put forward the working hypotheses on the anaphoric system of Spanish and I will provide an analysis that accounts for the semantic and syntactic properties of the inherent and non-inherent reflexive verbs, as well as the semantic differences introduced by the SE- and SELF-anaphors. Also the logophoric function of focus of the SELF-anaphors in Spanish will be briefly discussed. In section 4.4 I will defend that the differences between English, Dutch and Spanish reflexive verbs are due to mechanisms at the syntax-phonology interface. Finally, I will present the conclusions in section 4.5.

5.1. Introduction: Reflexivity in Romance and Germanic

In this section I will review the basic empirical data regarding reflexivization in English and Dutch (subsection 4.1.1), as well as in Spanish (subsection 4.1.2). Other languages such as German, French and Italian, may be sometimes mentioned here but are, nevertheless, left aside and out of the scope of this work.

5.1.1. Reflexivity in English and Dutch

As we saw in chapter 3 section 3.3.1.3., the Canonical Binding Conditions (CBC) in (1) are proposed by Chomsky (1981) in the framework of Government & Binding to account for the syntactic characteristics and the referential interpretation of the pronouns.

(1) Canonical Binding Conditions (CBC): (Chomsky 1981:188,211)

A. An anaphor is bound in its governing category.

B. A pronominal is free in its governing category.

where γ is a governing category for β iff γ is the minimal category containing β , a governor of β , and a SUBJECT (accessible to β); and α binds β iff α and β are coindexed and α c-commands β .

The CBC distinguish two types of pronouns: anaphors and pronominals. An anaphor is *himself* in (2a), which must be bound in its local configuration. This is defined by the concept *governing category*. A pronominal is *him* in (2b), which must be free (not bound by any antecedent) in its governing category. An antecedent outside the governing category can bind a pronominal though.

(2) a. Gandalf_i bewitched **himself**_{i/*j}.

b. Gandalf_i bewitched **him**_{j/*i}.

The CBC and the data in (2) summarize fairly well the basic facts concerning reflexivization in English¹. Moreover, Everaert (1986), Reinhart & Reuland (1991), Reuland & Koster (1991) and Reinhart & Reuland (1993) (among many others linguists) noted that the CBC were too restrictive so as to account for the behaviour of the anaphors in languages other than English. It can be seen in (3) that Dutch has a two-way anaphors (both (3a) and (3b)) unlike English, which has just one type of anaphor (2a).

¹ Although see more data that seem to contradict the facts described here further below in section 5.2.1.

- (3) a. Frodo waste **zich**
 Frodo washed **zich**
"Frodo washed (himself)"
- b. Frodo zag **zichzelf**
 Frodo saw **zichzelf**
"Frodo saw himself"

The anaphor *zichzelf* in (3b) is basically equivalent to *himself* in (2a), and it obeys Condition A. This kind of anaphors are SELF-anaphors (see section 5.2.1 for more details). On the other hand, the anaphor *zich* in (3a) does not obey Condition A (unlike *zichzelf*). Therefore *zich* can be bound by an antecedent outside its governing category as in (4). In this case, it can also alternate with the pronominal *hem* (similar to English *him*), which obeys Condition B. This kind of anaphors are SE-anaphors (see section 5.2.1 for more details).

- (4) Frodo_i zag [jou achter **zich**_i / hem_i staan]
 Frodo saw [you behind **zich** / him stand]
- (5) Smeagol_i haat ***zich**_i / zichzelf_i
 Smeagol hates **zich** / zichzelf
"Smeagol hates himself"

Another difference is that *zichzelf* (but not *zich*) is able to license a reflexive reading with verbs that are not marked as reflexives in the lexicon. In (5) only *zichzelf* can appear since the sentence has a reflexive interpretation and the verb *haten* (hate) is not lexically marked as reflexive.

Note in (6) that English uses zero-morphology with verbs like *wassen* (wash) in (3a) and (6a) instead of a SE-anaphor (though see section 5.4). With this kind of verbs, also a SELF-anaphor can appear (6b), but this is optional both in English and in Dutch. We will argue later on (section 5.3) that the difference between (3a) and (3b) is that in the former the verb is inherently reflexive (IRV) whereas in the latter the verb is non-inherently reflexive (nIRV).

- (6) a. Frodo washed Ø (himself).
 b. Frodo waste **zich** / **zichzelf**.
 Frodo washed **zich** / **zichzelf**
"Frodo washed (himself)"

5.1.2. Reflexivity in Spanish

Reflexivization in Spanish seems different from reflexivization in English and Dutch. There are three ways of marking reflexivization in Spanish (7):

(7) Reflexivization marks in Spanish:

- | | |
|--|--------------------------------|
| a. Clitics: | me, te, se, nos, os |
| b. (Morphologically) complex anaphors ² : | sí / mí / ti / él / etc.+mismo |
| c. (Morphologically) simple anaphors ³ : | si / mí / ti |

Simple anaphors in Germanic are not subject to Condition A (CBC) unlike complex anaphors (see chapter 3, section 3.3.2.3.). Nevertheless both kinds of anaphors seem to be subject to Condition A in Spanish. Note that even the sentence (9) below is a case of local binding since the PP *por sí y ante sí* is adjoined to the verbal phrase and c-commanded by the subject *Juan* (examples from Otero (1999)⁴).

- (8) *Ana_i le dijo a Luis_j [que Juan_k habló mal de sí_i]. Otero (1999:1437)

Ana him_{clitic} said to Luis [that Juan spoke badly of sí]

"Ana told Luis that Juan spoke badly about him."

- (9) Juan_i decidió [que Ana_j se hiciera cargo de la fábrica que hasta entonces había dirigido Luisa_k] por sí_i y ante sí_i. Otero (1999:1437)

Juan decided [that Ana se took management of the factory that till then had lead Luisa] by sí and before sí

"Juan decided himself that Ana took management of the factory that Luisa had lead until then."

Besides the examples above, in (10) and (11), it can be seen that *sí* obeys condition A (CBC) and hence, it has to be bound by the most local suitable antecedent⁵.

- (10) a. Rosa_i apretaba a Luis contra sí_i. Otero (1999:1445)

Rosa squeezed to Luis against sí

"Rosa squeezed Luis (against herself)."

² Otero (1999:1436) calls them *anáforas complejas* (complex) or *fuertes* (strong).

³ Otero (1999:1436) calls them *anáforas simples* (simple), *ligeras* (light) or *débiles* (weak).

⁴ The glosses and translations of Otero's (1999) examples are my own.

⁵ Note in (10) that *Luis* cannot bind *sí* because that would trigger a reflexive reading, and such a reading has to be licensed by a SELF-anaphor (*sí mismo*).

b. En las dificultades, el clan_i se plegaba siempre sobre sí_i bajo un espeso manto de silencio. Otero (1999:1445)

In the difficulties, the clan_i se bent always over sí_i under a dense cover of silence.

"In the difficulties, the clan always bent over itself under a dense cover of silence."

(11) a. *Rosa_i observó [que Tomasa apretaba a Luis contra sí_i]. Otero (1999:1445)

Rosa observed [that Tomasa huddled to Luis against sí.

"Rosa observed that Tomasa huddled Luis against her."

b. *El clan_i partía de la base de [que, en las dificultades, la familia se plegaba siempre sobre sí_i bajo un espeso manto de silencio]. Otero (1999:1445)

The clan stated of the base of [that, in the difficulties, the family se bent always over sí under a dense cover of silence.

"The clan stated out from the idea that, in the difficulties, the family always bent over itself under a dense cover of silence."

Further evidence of the fact that *sí* has to be locally bound in Spanish can be seen in sentences (12)-(17) below:

(12) a. Juan_i confía en sí_i. Otero (1999:1446)

Juan trust in sí

"Juan trust himself."

b. Juan_i insiste en [que Ana_j confía en sí_i/*i/j]. Otero (1999:1446)

Juan insists in [that Ana trust in sí]

*"Juan insists that Ana trusts *him/herself."*

(13) a. Juan_i no tenía confianza en sí_i. Otero (1999:1446)

Juan not had confidence in sí

"Juan did not trust himself."

b. Juan_i creía [que Ana_j no tenía confianza en sí_i/*i/j]. Otero (1999:1446)

Juan thought [that Ana not had confidence in sí]

"Juan thought that Ana did not had any confidence in him."

(14) a. [Las historias de Blas_i sobre sí_i] son muy divertidas. Otero (1999:1446)

[The stories of Blas about sí] are very funny

"Blas' stories about himself are very funny."

- b. Blas_i encuentra divertidas [las historias de Ana_j sobre **sí**_{*i/j}]. *Otero (1999:1446)*
 Blas finds funny [the stories of Ana about **sí**]
*"Blas finds the stories of Ana about *him/herself very funny."*
- (15) a. *Luisa_i encuentra divertidas [las historias de Juana_j sobre **sí**_i]. *Otero (1999:1446)*
 Luisa finds funny [the stories of Juana about **sí**]
"Luisa finds funny the stories of Juana about herself."
- b. Luisa_i encuentra divertidas [las historias sobre **sí**_i (contadas por ella o por otros)] *Otero (1999:1446)*
 Luisa finds funny [the stories about **sí** (told by herself or by others)]
- (16) a. Olga_i está orgullosa de **sí**_i. *Otero (1999:1447)*
 Olga is proud of **sí**
"Olga is proud of herself."
- b. Olga_i está segura de [que Ana_j está orgullosa de **sí**_{*i/j}]. *Otero (1999:1447)*
 Olga is sure of [that Ana is proud of **sí**]
"Olga is sure that Ana is proud of herself."
- (17) a. Juan_i nunca habla mal de **sí**_i. *Otero (1999:1447)*
 Juan never speaks badly of **sí**
"Juan never speaks badly about himself."
- b. *La madre de Juan_i nunca habla mal de **sí**_i. *Otero (1999:1447)*
 The mother of Juan never talks badly of **sí**
"Juan's mother never talks badly about him (him=Juan)."
- c. La madre_i de Juan nunca habla mal de **sí**_i. *Otero (1999:1447)*
 The mother of Juan never talks badly of **sí**
"Juan's mother never talks badly about herself."

In (18) and (19) below we can see the differences in the binding domains between Dutch *zich* and Spanish *si*. Whereas the former can be bound in Reuland and Koster's (1991) domain 2 (see chapter 3 above), i.e. the subject position of the matrix clause, the latter can only be bound in domain 1, i.e. within its clause.

- (18) a. *Rosa_i me hace [(a mí) trabajar para **sí**_i]. *Otero (1999:1438)*
 Rosa me_{clitic} makes [(to me) work for **sí**]
"Rosa makes me work for her."

b. Rosa_i laat me [voor zich_i werken]. Otero (1999:1438)

Rosa makes me [for **zich** work]

"Rosa makes me work for her."

(19) a. *Juan_i vio (a) [sí_i caer]. Otero (1999:1439)

Juan saw (to) [sí fall]

"Juan saw himself fall/falling."

b. Juan_i zag [**zich**_i vallen]. Otero (1999:1439)

Juan saw [**zich** fall]

"Juan saw himself fall/falling."

In (20) below we can see the differences in binding domain of Spanish *si*, which can only be locally bound in domain 1, and Italian *se*, which can be also be bound in domain 3 (examples from Otero (1999)):

(20) a. Aquel dictador_i pensaba que el pueblo_j hubiera sido mucho más feliz si los libros de historia hubiesen hablado más de sí_{i/*j} y de sus hazañas.

Otero (1999:1439)

That dictator thought that the folk had been much more happy if the books of history had talked more about sí and of his exploits.

"That dictator thought that the folk would have been happier if the History books had talked more about him and his exploits."

b. Quel dittatore_i pensava che ik popolo_j sarebbe stato molto più felice se i libri di storia avessero parlato di piú di sé_{i/*j} e delle sue gesta.

Otero (1999:1439)

That dictator thought that the folk had been much more happy if the books of history had talked more about sí and of his exploits.

"That dictator thought that the folk would have been happier if the History books had talked more about him and his exploits."

Sí cannot occupy nominative positions as seen in (21) and (22b), unlike *sich* in German (22a), nor accusative (23) positions. Hence, we can conclude that *sí* bears oblique Case⁶.

(21) - ¿Es que alguien fue a la fiesta en vez de Juan_i? (Otero 1999:1437)

Is that anyone went to the party instead of Juan?

⁶ Although see (26e) below.

- *No, es que **sí_i** fue. Otero (1999:1437)

No, is that **sí** went (sí = Juan)

(22) a. Wen wascht Otto? **Sich_i**. Otero (1999:1437; footnote 18)

Who washes Otto? **Sich** (sich = Otto)

b. *¿Quién lava a Otto? **Sí_i**. Otero (1999:1437; footnote 18)

Who washes to Otto? **Sí** (sí = Otto)

(23) *Anand_i parecía otra vez **sí_i** (mismo). Otero (1999:1437)

Anand looked again **sí** (self)

"Anand looked himself again."

Moreover, observe in (24) that *si* cannot refer to either the speaker (1st person pronouns) or the interlocutor (2nd person pronouns):

(24) a. *Yo_i estoy durmiendo en la habitación de **sí_i**. Otero (1999:1438)

I am sleeping in the room of **sí** (sí = I)

b. *Vosotros_i estáis durmiendo en la habitación de **sí_i**. Otero (1999:1438)

You_{plural} are sleeping in the room of **sí** (sí = you_{plural})

In (25) below we can see that *si* also shows characteristics that are not typical of what Reinhart & Reuland (1993) call *logophors*: (25a) shows that *si* can have an inanimate antecedent, and (25b) shows that *si* can refer to an antecedent that is not subject-oriented.

(25) a. Este salario_i no da mucho de **sí_i**. Otero (1999:1438)

This salary no gives much of **sí** (sí = itself)

b. Los piratas le dieron a Juan_i el tesoro para **sí_i**. Otero (1999:1438)

The pirates him_{clitic} gave to Juan the treasure for **sí** (sí = Juan)

Finally, Otero characterizes Spanish *si* as follows:

(26) Properties of *si*: Otero (1999:1440)

- a. *Sí* cannot refer to an antecedent in the discourse, as seen in (21) and (22b).
- b. *Sí* cannot occupy a nominative position, as seen in (21) and (22b).
- c. In general, *si* cannot refer to a 1st or 2nd person antecedent as in (24).
- d. *Sí* cannot be mid-distance bound unlike other pronominal elements such as Dutch *zich* (18) and (19). Neither can it be long-distance bound like Italian *se* (20).

- e. It always follows a preposition: i.e. *sí* bears oblique Case with few exceptions as in (32) below.
- f. *Sí* needs not be subject oriented as in (25b), and can refer to an inanimate antecedent as in (25a).

Sometimes, the presence of a clitic (7a) is enough so as to get a reflexive reading as in (27a). Other times, it is necessary the use of a morphologically complex anaphor (7b), as it happens in (27b). Finally, it can be seen that the morphologically simple anaphor *sí* in Spanish appears after a preposition as in (27c).

- (27) a. Juan **se** lavó
 Juan **se** washed
"Juan washed"
- b. Juan **se** besa **a sí mismo**
 Juan **se** kisses **to si mismo**
"Juan kisses himself"
- c. Ana_i vio una araña ante **sí_i**
 Ana saw a spider before **si**
"Ana saw a spider before her"

With inherent reflexive verbs (Otero 1999, Doron & Rapaport-Hovav 2007), a clitic (7a) is enough in order to license a reflexive reading of the predicate. These verbs can also appear with complex anaphors, though. In this case, there is an emphatic nuance, as in (28).

- (28) a. Juan *(**se**) lavó (**a sí mismo**)
 Juan (**se**) washed (**to himself**)
- b. Juan *(**se**) peina (**a sí mismo**)
 Juan (**se**) combs (**to himself**)

The nIRVs require a complex anaphor (7b) so as to get the reflexive reading as in (29).

- (29) a. María **se** criticó *(**a sí misma**)⁷
 María **se** criticized (**to herself**)
"María criticized herself."

⁷ Although see Torrego's (1995:229) example (14).

b. María **se** hace cosquillas **?(a sí misma)**

María **se** makes tickles **(to herself)**

"María tickles herself."

The complex anaphor has to be duplicated by a clitic when it occupies an argumental position where accusative or dative is available (Torrego 1995), as in (30). However, see the example (31) from "El Quijote" and the sentences in (32) of fixed expressions where the clitic needs not be doubled.

(30) María_i ***(se)** mira **a sí misma** (Torrego 1995:223)

Maria **(se)** sees **to herself**

"María sees herself"

(31) a. Yo he tomado el pulso a **mí mismo**. (Quijote II, iv; cited in Otero 1999:1458)

I have taken the pulse to **mí self**

"I have taken my pulse."

b. Yo ***(me)** he tomado el pulso a **mí mismo**. (Otero 1999:1458)

I **(me_{clitic})** have taken the pulse to **mí self**

"I have taken my pulse."

(32) a. Blas_i **(se)** es fiel a **sí mismo**. (Otero 1999:1459)

Blas **(se)** is loyal to **sí self**

"Blas is loyal to himself."

b. Tú **(te)** eres fiel a **tí mismo**. (Otero 1999:1459)

You **(te)** are loyal to **tí self**

"You are loyal to yourself."

c. Yo **(me)** soy fiel a **mí mismo**. (Otero 1999:1459)

I **(me)** am loyal to **mí self**

"I am loyal to myself."

On the contrary, when the complex anaphor occupies non-argumental positions or, in other words, positions that are not marked with accusative or dative, the anaphor cannot be duplicated by the clitic (Torrego 1995), as in (33).

(33) El presidente_i ***(se)** desconfía de **sí mismo** (Torrego 1995:224)

The president **(se)** distrust of **himself**

"The president distrusts himself."

As seen in (33) above and right below, the clitic is not necessary when *sí* is within a prepositional phrase (it bears oblique Case) or within a coordinated structure (we will see in section 5.3.3 that this is due to the impossibility of forming chains). Examples from Otero (1999).

- (34) a. Ana escribió una carta [[a la humanidad] y [a **sí/ella misma**]].
(Otero 1999:1460)

Ana wrote a letter [[to the mankind] and [to **sí/her self**]]
"Ana wrote a letter to the mankind and herself."

- b. [Ana (le) escribió a Blas] y [*(se escribió) a **sí misma**]. (Otero 1999:1460)
[Ana him_{clitic} wrote to Blas] and [(se wrote) to **sí self**]
"Ana wrote to Blas and wrote to herself."

- c. Ana ?*(se) escribió una carta [[a **sí/ella misma**] y [a la humanidad]].
(Otero 1999:1460)

Ana (se) wrote a letter [[to **sí/her self**] and [to the mankind]]
"Ana wrote a letter to herself and the mankind."

- (35) Ana escribió una carta [[sobre la condición humana] y [sobre **sí/ella misma**]].
(Otero 1999:1460)

Ana wrote a letter [[about the condition human] and [about **sí/her self**]]
"Ana wrote a letter about the human condition and about herself."

In non-argumental positions, both a simple anaphor and a complex one can be used, as in (36).

- (36) María_i tiene ante **sí_i / sí_i misma** un problema difícil (Torrego 1995:223)
Maria has before **sí / herself** a problem difficult
"María has a difficult problem before her."

Note that in (37b) the complex anaphor is subject to Condition A (CBC), and the pronominals like *ella* in (37a) are subject to Condition B (CBC). This is basically the same pattern followed by the SELF-anaphors and the pronominals in English.

- (37) a. *María_i se critica a **ella_i**⁸
María se criticizes to **her**
"María criticizes herself."

⁸ Although see Torrego's (1995:229) example (14).

b. María_i **se** critica a **ella_{i/*j}** **misma**

María **se** criticizes to **herself**

"María criticizes herself."

5.1.3. Recapitulation

We have seen that English uses zero morphology with inherent reflexive verbs and SELF-anaphors with non-inherent reflexive verbs (and optionally with IRVs). The CBC of Chomsky's (1981) were formulated in order to account for the distribution of SELF-anaphors and pronominals in English. However, other languages, such as Dutch, make use of a two-way anaphors, i.e. SE- and SELF-anaphors. Dutch uses SE-anaphors like *zich* with IRVs, whereas SELF-anaphors like *zichzelf* are require for nIRVs and optional with IRVs.

Spanish, on the other hand, has three marks of reflexivization: clitics with IRVs, morphologically complex anaphors with nIRV and optional with IRVs, and morphologically simple anaphors with nIRV that (almost) always require a preposition and thus, oblique Case.

In table (38) we can see a summary of the properties of English, Dutch and Spanish. In what rests of the chapter, I will give a unitary analysis of reflexivization in these three languages.

(38) Reflexivization in English, Dutch and Spanish:

ENGLISH	DUTCH	SPANISH	
Zero-morphology	SE-anaphor (<i>zich</i>)	Clitics (<i>me, te, se, nos, os</i>)	(IRVs)
SELF-anaphor (<i>himself</i>)	SELF-anaphor (<i>zichzelf</i>)	Complex anaphor (<i>sí/mi/ti+mismo</i>) (<i>él/ella/yo+mismo</i>)	(nIRVs) (optionally IRVs)
		Simple anaphor (<i>mí, tí, sí</i>)	(nIRVs + PP)

5.2. Theoretical background: The lexicon-syntax interface and reflexivization

This section reviews (and in some cases revises) the theoretical background upon which the analysis of reflexivization to be developed in sections 5.3 and 5.4 is based.

Firstly, I will go a step further from Chomsky's (1981) Canonical Binding Conditions in section 5.2.1, and review Reinhart & Reuland's (1993) Conditions A and B on reflexivization.

Secondly, I will follow the model proposed by Reinhart (2002) for the lexicon-syntax interface: the Theta-System. This will be understood as a system that interfaces between the system of concepts and the different linguistic systems (C-I, C_{HL}, S-M), and comprises not only a lexicon but also a set of lexical operations. Hence, its conception as an "active lexicon" (Siloni 2002, Reinhart & Siloni 2005).

Thirdly, I will review the null SE-anaphor PRO analysed in chapter 4 and revise our ontology of null SE-anaphors in the Universal Grammar so that it can account for the reflexivization in the languages under study (English, Spanish and Dutch).

Finally, I will recapitulate and summarize the theoretical tools I will make use of in this chapter and the rests of the thesis. I will distinguish two strategies to get reflexive readings: lexical reflexivization and reflexive binding.

5.2.1. Beyond the Canonical Binding Conditions: SE- and SELF-anaphors

The CBC predict strict complementary distribution of anaphors and pronominals. In other words, whenever an anaphor (subject to Condition A) can appear, a pronominal (subject to Condition B) cannot appear, as in (39a). However, Reinhart & Reuland (1993) provided empirical evidence against this prediction. Sentences can be found where both an anaphor and a pronominal are allowed to appear, as in (39b,c).

- (39) a. The queen invited [**me** / ***myself**] for tea. (Reinhart & Reuland [R&R⁹] 1993:675)
 b. The queen invited both Max and [**myself** / **me**] for tea. (R&R 1993:675)
 c. Max_i said that the queen invited both Lucie and [**himself_i** / **him_i**] for tea.
 (R&R 1993:675)

Myself in (39b) and *himself* in (39c) are SELF-anaphors that violate Condition A (CBC), i.e. they do not have an antecedent in the same sentence that binds them within their local domain. These anaphors are considered *logophors*, i.e. anaphors with a special discursive function. Reuland & Koster (1991), Reinhart & Reuland (1991, 1993) and Reuland (2006b, 2006c) pay attention to this phenomenon so as to integrate the discursive factors within the binding theory. Reinhart & Reuland (1993) argue that *myself* in (39b) has an emphatic function, whereas *himself* in (39c) expresses the source

⁹ I will use "R&R" hereinafter to refer to "Reinhart & Reuland".

of the discourse, i.e. refers to the individual that said something, *Max* in this case (see also Reuland 2006b).

Reinhart & Reuland (1993) argue that the binding theory should be stated on the basis of the formation of reflexive predicates rather than on nominal binding. They reformulate in (40) the binding conditions in terms of conditions on reflexivization. Furthermore, these conditions do not apply on nominal elements (pronominals and anaphors) but on predicates (syntactic and semantic predicates), defined in (41).

(40) Reinhart & Reuland's Conditions (1993): (R&R 1993:678)

A. A reflexive-marked predicate is reflexive.

B. A reflexive predicate is reflexive-marked.

(41) Definitions (Reinhart y Reuland 1993): (R&R 1993:678)

a. The syntactic predicate formed of (a head) P is P, all its syntactic arguments, and an external argument of P (subject). The syntactic arguments of P are the projections assigned θ -role or Case by P.

b. The semantic predicate formed of P is P and all its arguments at the relevant semantic level.

c. A predicate is reflexive iff two of its arguments are coindexed.

d. A predicate (formed of P) is reflexive-marked iff either P is lexically reflexive or one of P's arguments is a SELF-anaphor.

The conditions defined in (40) account for the cases in (39b,c) that escape the CBC of Chomsky. In (39a), the predicate cannot have a reflexive reading. Therefore a SELF-anaphor that marks the predicate as reflexive would violate R&R's Condition A¹⁰. In (39b) the predicate is not reflexive, but *myself* does not mark the syntactic predicate as reflexive since it is part of a coordination structure. Therefore, its presence does not violate R&R's Condition A. *Himself* in (39c) is part of a coordinate structure and therefore cannot mark the predicate as reflexive. Hence, R&R's Condition A is not violated.

Reinhart & Reuland identify predicative heads (verbs) that are lexically reflexives (41d), which we will call *inherent reflexive verbs* (IRVs). If a predicative head is not an IRV (i.e. it is a *non-inherent reflexive verb* or nIRV), a SELF-anaphor is necessary so as

¹⁰ I will use "R&R" to differentiate Reinhart & Reuland's (1993) Conditions from Chomsky's (1981) Conditions, for which "CBC" will be used.

to license the reflexive reading of the predicate and to fulfil R&R's Condition B. We will come back to this issue later in section 5.2.4.

Anaphors like *zich* in Dutch, which do not obey Condition A (neither do they obey Condition B¹¹), have traditionally been called *long distance anaphors*, *pronominal anaphors*, *simple anaphors* or *SE-anaphors* (term used by Reinhart & Reuland 1993, among others). Examples of SE-anaphors are *zich* in Dutch (Reinhart & Reuland 1993), *sig* in Icelandic (Reuland 2001) and, as we claimed in chapter 4, also PRO. These anaphors are simpler in their morphological expression (they are usually deficient in ϕ -features) than complex anaphors like English *himself* or Dutch *zichzelf*. The latter are usually¹² subject to Condition A. Another difference is that SE-anaphors, unlike SELF-anaphors, cannot license a reflexive reading with a verb that is not lexically marked as reflexive, as seen in (5) and repeated below in (42).

- (42) Smeagol_i haat ***zich**_i / **zichzelf**_i
 Smeagol hates **zich** / **zichzelf**
 "Smeagol hates himself"

Based on these properties, Reinhart & Reuland (1993) proposed the following classification of lexical anaphoric expressions:

- a) **Pronouns** are subject to Condition B (CBC), cannot legitimate reflexive readings with verbs that are not inherent reflexives, and have a [+R]¹³ feature, which implies that they can directly refer to an entity in the discourse, i.e. they can get a value directly from the discursive context.
- b) **Anaphors**, which can be classified in:
 - i. **SE-anaphors** are not subject to Condition A or Condition B (CBC). They cannot license reflexive readings when the verb is not marked as reflexive in the lexicon, and they are [-R]¹⁴. They are elements such as *zich* in Dutch or

¹¹ Reuland (2001) presents some examples where the SE-anaphor *zich* is locally bound in Dutch.

¹² In certain syntactic contexts, the SELF-anaphors can violate Condition A (CBC) and license a logophoric reading (section 5.3.5).

¹³ Referential independence (Reinhart & Reuland 1993).

¹⁴ PRO can indeed get a value from the discourse storage in non-obligatory control environments. Hence I state that PRO is [+R]. I propose that this is the great difference between the spelled-out SE-anaphors like

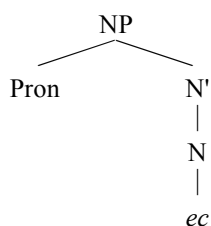
sig in Icelandic (Reinhart & Reuland 1993, Reuland 2001), *se* in Spanish (see section 5.3) and PRO (see chapter 3).

- ii. **SELF-anaphors** are subject to Condition A (CBC), are [-R] and can license reflexive readings with non-inherent reflexive verbs. Instances of this type of anaphor are *himself* in English, *zichzelf* in Dutch (Reinhart & Reuland 1993, Reuland 2001). In section 5.2 I will argue that the Spanish complex anaphors that follow the pattern $x+mismo$ are SELF-anaphors too.

The structure of the pronouns is schematized in (43), whereas (44) and (45) represent the structure of the SE- and the SELF-anaphors respectively. The pronouns and the SE-element are in the determiner position, though they project a NP (Reinhart & Reuland 1993). The SELF- element is a nominal element with a pronoun or SE- element that occupies the determiner position (though see Reuland 2009 for German). Note that the pronouns have a full set of ϕ -features, whereas the SE-anaphors have a defective set of ϕ -features.

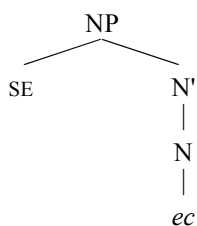
(43) Pronominals:

(full set of ϕ -features; R&R 1993:658)



(44) SE-anaphors:

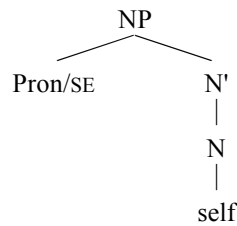
(defective set of ϕ -features; R&R 1993:658)



zich and PRO. The latter is [+R] due to its ϕ -feature grammatical number (despite its being unvalued). *Zich*, on the contrary, seems not to have a grammatical number feature, and hence it is [-R]. Despite the fact that PRO has person and number feature, it can nevertheless form chains due to its lack of gender feature. The full discussion of this issue is beyond the scope of this work, however. See Teomiro (in progress) for a more extended discussion on this issue.

(45) SELF-anaphors:

(R&R 1993:658)



The table in (46) summarizes the properties of the three types of anaphoric expressions previously described, according to Reinhart & Reuland. Later on, this table will be extended so as to include other pronominal elements (such as PRO and the Spanish clitic *se*).

(46) Typology of anaphoric expressions:

(R&R 1993:659)

	SELF	SE	PRONOUN
REFLEXIVIZING FUNCTION	+	-	-
R (REFERENTIAL INDEPENDENCE)	-	-	+

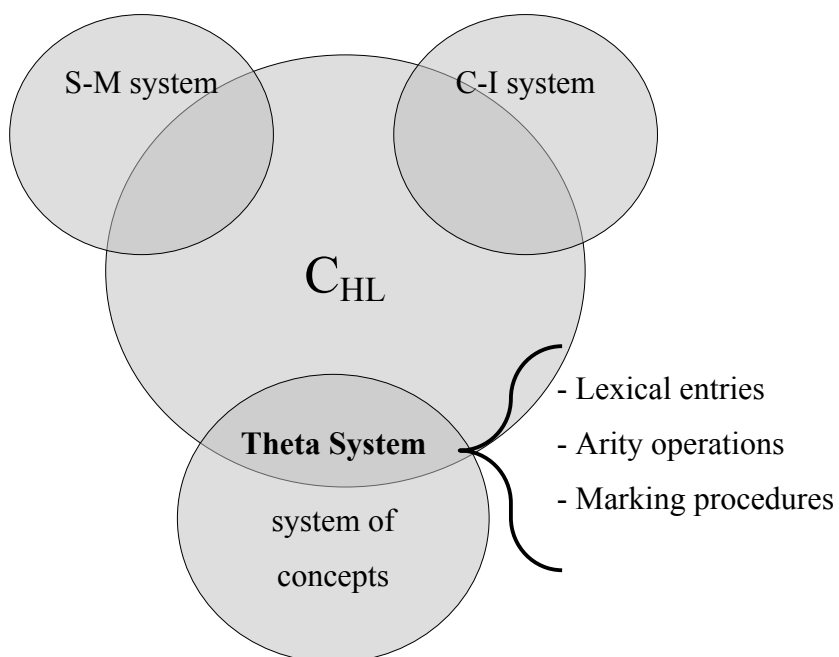
5.2.2. Active lexicon and the Theta System

In order to study reflexivity, we need to take into account thematic relations. More generally, we need a model of how the system of concepts, the lexicon and the syntax interact with one another. In other words, we need a theoretical model of the lexicon-syntax interface. I will follow Reinhart's (2002) model, which she calls the Theta System (Reinhart 2002) and defines as the system that interfaces between the system of concepts and the computational system of human language (C_{HL}), as graphically represented in (48). It consists of (at least) the elements in (47):

(47) The Theta System consist of (at least):

(Reinhart 2002:229-230)

- a. **Lexical entries**, which are coded concepts. Very little semantic information of the concepts is visible to the syntax, and this information is codified by formal features defining the thematic relations of verb and its arguments. The lexicon has traditionally been understood as the collection of these lexical entries.
- b. **Arity operations** on lexical entries, which may generate new entries, or just new options of realization.
- c. **Marking procedures**, which prepare a verb entry for syntactic derivation.

(48) The Theta System:*(based on Reinhart 2002)*5.2.2.1. *Lexical entries, θ -features and θ -clusters*

I will also follow the proposal of Reinhart (also of authors like Hornstein (1999) and Reinhart & Siloni (2005)) on theta features: the theta (thematic) relations between the predicate and its arguments are encoded in the syntax by θ -features¹⁵, which in the Theta System are those in (49)

(49) θ -Features:*(Reinhart 2002:230-231)*

- a. [**c**]: is associated with an argument that is perceived as a sufficient condition for the action described by the verb.
- b. [**m**]: is associated with some sort of mental state of the participant though it does not determine the causal status of the argument (i.e. whether or not it is a sufficient condition).

These two features combine so as to form eight θ -clusters, which can be formed out of one or two θ -features. The θ -clusters are specified in (50), and basically correspond to the thematic roles identified in the traditional literature on theta roles and thematic relations. A θ -cluster can be positive if the two θ -features that contains are positive, as

¹⁵ In other words, the claim that I follow in this chapter is that the θ -roles are encoded by means of features rather than assigned by virtue of syntactic relations, as Chomsky (1981), among many other authors, has argued.

in (50a, e, g). It is negative if its two θ -features are negative, as in (50c,f,h). When a feature is positive, and the other is negative, the cluster is mixed, as in (50b,d).

(50) θ -clusters: (Reinhart 2002:232)

a. [+c+m] Agent	e. [+c] Cause
b. [+c-m] Instrument	f. [-c] Goal/Benefactor
c. [-c-m] Theme/Patient	g. [+m] Sentient
d. [-c+m] Experiencer	h. [-m] Subject matter
	i. [] Undefined ¹⁶

The clusters agent, instrument, theme and experiencer are fully specified for θ -features, i.e. they have both θ -features specified. The other clusters (cause, goal, sentient and subject matter) are undefined for one of the two θ -features. This implies that they have more interpretative freedom. For example, a cause cluster [+c] can be interpreted either as an agent [+c+m] or as an instrument [+c-m] depending on the contextual information. In other words, the context determines the value of the undefined θ -features, as can be seen in (51):

- (51) a. break V_{acc} ([+c]₁, [-c-m]₂)
 b. Mary_[+c+m] broke the window_[-c-m]
 c. The stone_[+c-m] broke the window_[-c-m]

The lexical entry of the verb *break* (51a) contains two θ -clusters: a fully defined [-c-m] theme cluster, and a unitary [+c] cluster. The unitary cluster has to be fully defined at C-I in order to be interpreted (Marelj 2004). The [+c] cluster is "expanded" according to semantic and pragmatic reasons. In (51b) the argument *Mary* is mentally involved (and thus, *Mary* is animated and human, otherwise it would be difficult for her to be mentally involved). Hence, it is interpreted as an agent [+c+m]. In (51c) the argument *stone* is non-animated and non-human, hence it is interpreted as an instrument or cause [+c-m].

5.2.2.2. Lexical marking and syntactic merging

The theta clusters are marked in the lexicon with *merging indexes* so as to be processed by the syntax (C_{HL}). The marking procedures and the indexes are specified in (52). A negative cluster (almost always) receives an index 2 (52a), whereas a positive cluster (almost always) receives an index 1 (52b). Mixed clusters remain unmarked. The

¹⁶ See Marelj (2004).

indexes 1 and 2 determine how the arguments merge in the syntactic derivation. The merging instructions in (53) are the way in which the C_{HL} interprets the merging indexes: a cluster marked with an index 2 merges within the vP , whereas a cluster marked with an index 1 merges out of the VP (more concretely, in $[Spec, vP]$). Mixed clusters (which remain unmarked) can merge either within the vP or out of VP (53a), depending on the syntactic configuration. For example, an instrument cluster $[+c-m]$ cannot merge out of the VP if there is an argument that realizes a cluster $[+c]_1$ that obligatorily merges out of the VP by (53b) (and there is only one external position), as in (54a). However, the instrument can, and has to merge out of the VP whenever there is no other argument marked with the index 1 due to (53a), as it happens in (54b).

(52) Lexicon marking: (Reinhart 2002:246)

Given an n -place verb entry, $n > 1$,

- a. mark a $[-]$ cluster with index 2
- b. mark a $[+]$ cluster with index 1

(53) C_{LH} merging instructions: (Reinhart 2002:247)

- a. when nothing rules this out, merge externally (out of the VP),
- b. an argument realizing a cluster marked 2 merges internally (within the vP), an argument realizing a cluster marked 1 merges externally (out of the VP).

- (54) a. Mary $[+c]_1$ broke the window with the stone $[+c-m]$.
 b. The stone $[-c+m]$ broke the window.

5.2.2.3. θ -Chains

I propose that the fact that a lexical verbal entry is specified with θ -clusters be understood in the following way: the lexical verbal entries (the V heads) have one or more θ -clusters that are valued on V but uninterpretable on V. Each cluster is marked with merging indexes, which are formal features that indicate to the C_{HL} how the argument must merge.

(55) $V [u \theta [\alpha C, \beta M]_{1/2/\text{none}}]$

On the other hand, the nouns (the N heads) have one unvalued and interpretable θ -cluster (i.e. the θ -features are interpreted on the nominals). The value of the θ -cluster is assigned by an agree operation with the valued instance of the θ -cluster on the verb.

(56) $N [i \theta [\text{unvalued}]]$

This means that the values of the thematic relations are introduced in the syntactic derivation by the verbs (V), and the nominals (N) get these values by means of agree operation. The thematic relations are interpreted on the nominals, since they express information on the relation of the argument with the event described by the verb.

The thematic assignment is realized by the agree operation, which forms agree-chains. These chains make it possible to share one θ -cluster by a verb and a nominal item.

(57) θ -chain:

- a. $\alpha [i \theta [-]] \& \beta [u \theta [val]] \rightarrow Agree \rightarrow \alpha [i \theta [val]] \& \beta [u \theta [val]]$
- b. John_[+c+m] reads.
- c. John_{[i θ [-]]} & reads_{[u θ [+c+m]]} $\rightarrow Agree \rightarrow$ John_{[i θ [+c+m]]} & reads_{[u θ [+c+m]]}
- d. θ -chain { John, reads }

Finally, we can extend the typology of agree-chains defined in chapter 2 (section 2.3) and 2 (section 2.3.4) with this last kind of chain:

(58) Types of agree-chains (revisited):

- a. **ϕ -chain:** agree-chain formed when two or more lexical items share one or more ϕ -features (see chapter 3, section 3.3.4.).
- b. **Tns-chain:** agree-chain formed when two or more lexical items share a Tns (tense) feature (see chapter 3, section 3.3.4.).
- c. **θ -chain:** agree-chain formed when two or more lexical items share a θ -feature as in (57).

To conclude, when a verb θ -marks a nominal element, this marking is formalized by the notion of θ -chain. In other words, a nominal element is θ -marked when it forms part of a θ -chain with a verbal head.

5.2.2.4. *Active lexicon and arity operations I: Reflexivization*

Some entries of the verb are listed in the lexicon as such, like (59a,d,f); this is to say that they are basic entries in the lexicon. Some others are derived from the basic ones by means of a lexical operation. The unaccusative verb in (59b) is considered to be derived from the transitive entry in (59a) (cf. Reinhart 2002 and Reinhart & Siloni 2005; although see other proposals in Pesetsky 1995 and Hale & Keyser 1997). So is the

passive verb in (59c). We will see that the reflexive entry of (59e) is derived from the transitive agentive entry in (59d). And finally, the transitive entry of verbs like *walk* in (59g) is thought to be derived from an unergative entry of *walk* like that in (59f).

- (59) a. John_[+c] / the wind_[+c] broke the window_[-c-m]. (transitive causative)
 b. The window_[-c-m] broke. (intransitive unaccusative)
 c. The window_[-c-m] was broken (by John_[+c]). (passive)
 d. John_[+c+m] washed Mary_[-c-m]. (transitive agentive)
 e. Mary_{[+c+m]+[-c-m]} washed. (reflexive)
 f. John_[+c+m] walks in the park every day. (intransitive unergative)
 g. John_[+c+m] walks the dog_[-c-m] in the park every day. (causativized transitive)

There are three types of operations that can apply to the verb's theta grid: *saturation*, *reduction* and *expansion*. Saturation applies in the formation of passives like (59c). Reduction reduces the verb's arity by one. The outcomes of reduction operations are unaccusative verbs as (59b) (we call this operation *decausativization*), and reflexive verbs as (59e) (we call this operation *reflexivization*). Expansion augments the verb's arity in one as we see in (59g).

We will see only one reduction operation in this chapter: reflexivization. This is the operation that forms IRVs like *washed* in (59e), with which we are concerned in this chapter. The other reduction operation, *decausativization*, as well as the expansion operation *causativization* will be seen in chapter 6. For a more detailed explanation on such arity operations see Chierchia (1989, 2004); Reinhart (2002); Reinhart & Siloni (2005) and Hovav & Siloni (2008a, 2008b).

Reflexivization is a lexicon operation that has been widely assumed derives a reflexive entry from a transitive one. I will follow Reinhart & Siloni's (2005) analysis and assume it operates as in (60).

- (60) Reflexivization: (based on Reinhart & Siloni 2005)
- a. Transitive (basic) entry: $V_{acc} ([+c+m]_1, [-c-m]_2)$
 b. Reflexivized entry: $R(V_{weak\ acc}) ([+c+m]+[-c-m])_1$
 c. Syntactic realization: $DP_{[+c+m]+[-c-m]1} V$
 d. Interpretation: $\exists e \lambda x [e=V \ \& \ [+c+m],e = x \ \& \ [-c-m],e = x]$

This operation applies to agentive transitive verbs, i.e. transitive verbs whose subjects are agents rather than causes. Therefore, this operation applies to verbs such as *wash*, *comb*, and *shave* (among many other) but not to verbs such as *break*, or *kill*.

The input for the operation to take place is a transitive agentive verb with an accusative feature. The operation consists of theta-bundling, i.e. the internal [-c-m] feature is bundle with the external [+c+m] feature in a compound theta-feature. The latter is marked with an index 1, by which the verb behaves as an unergative verb. The accusative feature is reduced to weak accusative or no accusative at all, depending on the language (see Reinhart & Siloni 2005). I will come back to this lexical operation later on in section 5.2.4.

The output of the reflexivization operation is a unergative verb. Kayne (1975) argued that French reflexives do not pattern with transitive verbs and hence, the reflexive clitic cannot be simply considered the object clitic of a transitive entry such as Burzio (1986) claimed for Italian object and reflexive clitics (see chapter 4 section 4.4.1.). Additional evidence against the object clitic analysis of reflexive verbs is suggested by Marantz (1984) and Sportiche (1998).

Other linguists have proposed that the reflexive clitic is associated with the external theta-role, and the reflexive verb is therefore an unaccusative verb, as its internal argument is the derived subject. Among the defenders of the unaccusative approach we can find Bouchard (1984), Grimshaw (1990), Marantz (1984), Kayne (1988), Pesetsky (1995) and Sportiche (1998). Mendikoetxea (1997) argues that reflexive verbs are actually transitive. They have a DP complement to the verb, and PRO in internal subject position. The clitic is the spell out of the feature person of AgrS, which allows the presence of PRO. However, and due to the inability of PRO to check the D-feature of AgrS, the object must raise to subject position.

Finally, other linguists have argued that reflexive verbs are (or behave as) unergative verbs such as Aranovich (2000) and Rodríguez Ramalle (2007). Reinhart & Siloni (2004, 2005), Reinhart (2002), argue against this theoretical approach to reflexive verbs and claim that the unaccusative analysis of the reflexive verbs must be discarded, as reflexives fail syntactic tests of unaccusativity:

- a. **Morphological arguments, such as auxiliary selection**, held by the defenders of the unaccusative hypothesis of the reflexive verbs seem not to be very strong. The fact that different diatheses of a verb may appear with the same morphological realization does not mean that their derivation are of the same nature, nor that they

share the same type of grammatical subject. Auxiliary selection is an intricate matter, which still poses a great number of problems for a proper analysis. Note furthermore that neither reflexives nor unaccusatives consistently choose the auxiliary *be* cross-linguistically.

- b. **En-cliticization in French:** the French quantitative clitic *en* can cliticize only out of the object position, thus it can serve as a diagnostics for unaccusativity as it discriminates between the internal and external argument in postverbal position. (61a) contains an unaccusative entry where *en*-cliticization is possible (62a). (61b) is a reflexive verb and disallows *en*-cliticization as can be seen in (62b). (61c) is an unaccusative with "reflexive morphology" (the clitic *se*) that allows *en*-cliticization as can be seen in (62c). This is straightforward as the subject of reflexives is an external argument, unlike the subject of unaccusatives.

(61) a. Il est arrivé trois filles hier soir. (Reinhart & Siloni 2004:172)
 there is arrived three girls yesterday evening
"There arrived three girls yesterday evening."

b. (?)Il s'est lavé beaucoup de touristes dans ces douches publiques, récemment.
 (Reinhart & Siloni 2004:172)
 there *se* is washed many tourists in these public showers recently
"Many tourist washed in this public showers recently."

c. Il s'est cassé beaucoup de verres dans ce lave-vaisselle.
 (Reinhart & Siloni 2004:172)
 there *se* is broken many glasses in this dish-washer
"Many dishes broke in this dishwasher."

(62) a. Il en est arrivé trois hier soir. (Reinhart & Siloni 2004:172)
 there of-them_{cl} is arrived three yesterday evening

b. *Il s'en est lavé beaucoup dans ces douches publiques, récemment.
 (Reinhart & Siloni 2004:172)
 there *se* of+them_{cl} is washed many in these public showers recently

c. Il s'en est cassé beaucoup dans ce lave-vaisselle.
 (Reinhart & Siloni 2004:172)
 there *se* of+them_{cl} is broken many in this dish-washer

c. **Ne-cliticization in Italian:** the same pattern holds in Italian, as shown in (63)^{17, 18}:

(63) a. Ne sono arrivati tre. (Reinhart & Siloni 2004:172)
 of+them_{cl} are arrived three

b. *Si ne sono vestiti tre. (Reinhart & Siloni 2004:172)
 si of+them_{cl} are dressed three

d. Even in English it seems that there is evidence that the subject of reflexives is an external argument. Agent nominals, also known as *-er* nominals can be derived only from predicates with an external argument. Hence, the contrast between (64a) and (64b). As expected, reflexives pattern with unergatives: they can give rise to agent nominals (64c).

(64) a. She runs so fast because she is an experienced runner. (Reinhart & Siloni 2004:175)

b. *She moves so gracefully because she is an experienced mover.
 (Reinhart & Siloni 2004:175)

c. She dresses slowly because she is an elegant dresser. (Reinhart & Siloni 2004:175)

Other tests used by Reinhart & Siloni (2004) are reduced relatives, modification by possessive datives, and genitive of negation in Russian.

Their claim is that reflexive verbs are unergative entries whose subject is an external argument unlike the subject of unaccusatives. Although the matter is not settled, I will pursue Reinhart & Siloni (2004, 2005) in that reflexive verbs are unergative entries.

5.2.3. Null SE-anaphors

I will adopt the analysis I elaborated in chapter 3 on PRO as a null SE-anaphor with interpretable and unvalued ϕ -features, and with an uninterpretable and unvalued Tns feature (structural Case). The feature composition of PRO is summarized in (65) below.

PRO is interpreted depending on the structural configuration, as we saw in chapter 4. When agree-chains can be formed between PRO and a suitable antecedent, PRO is interpreted as a bound anaphor, and obligatory control raises. When no agree-chains can

¹⁷ Examples cited by Reinhart & Siloni (2004), from Guglielmo Cinque (personal communication, cited by Grimshaw 1990:184n3).

¹⁸ Reinhart & Siloni (2004) note in their footnote 8 that some Italian speakers accept (63b), while others categorically rule it out.

be formed, PRO is interpreted as a pronominal, and non-obligatory control raises (see chapter 3 for a more detailed discussion).

As I argued in chapter 3 section 3.4., the introduction of PRO as a null SE-anaphor in the theoretical machinery is at no cost since we have empirical evidence on defective anaphors (Reinhart & Reuland 1993, Reuland 2001, among many others), as well as on the existence of null anaphors (Holmberg 2005).

I propose here that there is (at least) one other such null SE-anaphor available in UG, which is the least defined nominal item available in UG. Its feature composition is given in (65) below, and the difference with PRO is that it lacks grammatical number, which renders it [-R]. I will call this null SE-anaphor PRO'.

(65) Feature composition of null SE-anaphors:

	PRO ¹⁹	PRO'
ϕ -FEATURES	<i>i</i> ϕ person [unvalued] <i>i</i> ϕ number [unvalued]	<i>i</i> ϕ person [unvalued]
θ -FEATURE(S)	<i>i</i> θ [unvalued]	<i>i</i> θ [unvalued]
TNS-FEATURE (CASE)	<i>u</i> T [unvalued]	<i>u</i> T [unvalued]
PHONOLOGICAL CONTENT	no	no

Later on, I will argue that PRO' is inserted in the syntactic derivation whenever an arity reduction operation takes place at the lexicon in order to preserve the valence of the verb and the formal requirements of the syntax.

5.2.4. Recapitulation: Lexical reflexivization and reflexive binding

I will follow Doron & Rapaport-Hovav's (2007) proposal that IRVs are the result of a reflexive operation on theta roles in the lexicon (as in Reinhart 2002 and Reinhart & Siloni 2005). The nIRVs license a reflexive reading by means of a reflexive binding, i.e. a SELF-anaphor bound by an antecedent (Reuland 2001).

¹⁹ In Teomiro (2005) I defended that PRO has a full set of ϕ -features. This cannot be the case, however. If PRO had a full set of ϕ -features, it could not form a chain to be translated to A-binding at C-I. Note that Reuland (2001) stated as a necessary condition for chain formation the underspecification for some ϕ -feature. Therefore, PRO has to be underspecified for at least one ϕ -feature. The number feature is necessary since PRO is [+R] (as can be seen in non-obligatory control configurations), and person is needed in order to delete the uninterpretable instance on the verb. Therefore, gender is the most suitable candidate.

The IRVs are derived from a lexical operation of reflexivization (see section 5.2.2.4 above), by which the internal θ -cluster is reduced and bundled with the external θ -cluster. Moreover, the verb is no longer able to assign accusative Case (Reinhart and Siloni 2005). I will follow Reinhart and Siloni's formalization of the reflexivization operation in (66). Note that these verbs behave as unergative verbs in the syntax (according to Reinhart & Siloni 1999, 2005; Reinhart 2002). This points towards the possibility that they have just one syntactic argument, although the predicate receives a reflexive interpretation (with two semantic arguments) in the C-I system. I will add to Reinhart's formalization that despite the presence of two semantic arguments in the C-I system, these arguments are bound by just one lambda operator. This is because the reduction of the internal θ -cluster prevents the projection of a second lambda operator in the semantics.

- (66) Lexical reflexivization: *(based on Reinhart & Siloni 2005)*
- a. Transitive (basic) entry: $V_{acc} ([+c+m]_1, [-c-m]_2)$
 - b. Reflexivized entry: $R(V_{weak\ acc}) ([+c+m]+[-c-m])_1$
 - c. Syntactic realization: $DP_{[+c+m]+[-c-m]} V$
 - d. Interpretation: $\exists e \lambda x [e=V \ \& \ [+c+m], e = x \ \& \ [-c-m], e = x]$

The nIRVs enter in the syntactic derivation with their lexical entry unaltered. The reflexive interpretation comes from A-binding rather than a reflexivization operation in the lexicon.

- (67) Reflexive binding: *(based on Doron & Rappaport-Hovav 2007)*
- a. Lexical entry: $V_{acc} ([+c+m]_1, [-c-m]_2)$
 - b. Syntactic realization: $DP_{[+c+m]} V DP_{[-c-m]}^{SELF-anaphor}$
 - c. Interpretation: $\exists e \lambda x \lambda y [e=V \ \& \ [+c+m], e = x \ \& \ [-c-m], e = y \ \& \ y=f(x)]$

The nIRVs do not undergo any reflexivization operation in the lexicon. Therefore, their argumental structure requires two syntactic arguments. The SELF-anaphor makes it possible to bind a local antecedent without violation of the Thematic Criterion. This is due to the presence of the protector SELF-element (Reuland 2001).

Reuland shows that in Dutch there can be local binding between an antecedent and a SE-anaphor. However, this binding forces the two elements to be interpreted as just one element in the semantic system (because there is just one chain and the SE-anaphor is defective in ϕ -features). Verbs like *voelen* (feel), whose argumental structure

requires only one semantic argument, allows that *zich* be bound by the subject (there is no Theta Criterion violation).

(68) a. [Jan]_i voelt [**zich**]_i goed.

Jan feels **zich** well

(2 syntactic arguments, 1 semantic argument)

However, verbs like *haten* (hate) require two arguments in the syntax and in the semantics, since they are nIRVs that have not undergone any reduction operation in the lexicon. If there were not two arguments in the semantics, there would be a Theta Criterion violation. In order to prevent the anaphor and its antecedent from being interpreted as one semantic argument, Reuland argues that a protector SELF-element is added. Hence, the chain is not formed between the antecedent and the anaphor - the τ phrase in (69) - but it is formed between the antecedent and the SE-element (within the τ phrase) of the SELF-anaphor. This chain is interpreted as A-binding but there are two syntactic objects that are translated to two distinguishable semantic objects, since the antecedent and the τ phrase do not form a chain themselves. The chain is formed between *zich* (the SE-element) and the antecedent. Therefore, the reflexive binding does not violate the Theta Criterion.

(69) a. [Jan]_i haat [_i **zich** [_N **zelf**]]

Jan hates **zich** **zelf**

(2 syntactic arguments, 2 semantic arguments)

Note an important difference between the lexical reflexivization (66) and the reflexive binding (67). In the lexical reflexivization, both the subject and the object are exactly the same individual since there are two variables in the semantic representation bound by one lambda operator. On the other hand, in the reflexive binding there are two variables bound by two lambda operators. The SELF-element not only has a protective function in the syntax, but it also introduces an identity function in the semantic (70), which forces the second variable to be interpreted as a function of the first one. See Reuland (2001:481-486) for a more detailed explanation. See Otero (1999:1448-1459) on the interpretation of *MISMO* in Spanish.

(70) $x R f(x)$

(Reuland 2001:481)

In section 4.3 it will be shown how this function is responsible of the different semantic interpretation of the IRVs with SE-anaphors and the nIRVs that require SELF-anaphors in syntactic configurations where both kinds of anaphors can alternate.

5.3. SE-anaphors at the lexicon-syntax and syntax-semantics interfaces

Based on the data presented in section 5.1.2, I will put forward three hypotheses on the reflexivization and the anaphoric system of Spanish for the sake of clarity, which will be applied and contrasted in the subsequent sections²⁰:

- A) **Spanish has SELF-anaphors²¹**: they are the anaphors in (7b), and follow the pattern $x+mismo$. They are [-R], are subject to Condition A (CBC) and can license reflexive predicates when the verb is a nIRV (their presence is enough to fulfill R&R's Condition B). There is covert movement of the element MISMO to the predicative head (Reinhart & Reuland 1993). When the SELF-anaphor does not mark a predicate as reflexive, it can be used as a logophor (see section 5.3.5).
- B) **Spanish has SE-anaphors too²²**: besides PRO (chapter 5), there are other SE-anaphors in Spanish that are [-R] and cannot license a reflexive reading by themselves (i.e. their presence is not enough for a nIRV to fulfill R&R's Condition B). They are not subject to Condition A or B (CBC), and they can be divided in two types:
- a. **non-tonic**: the clitics *me, te, se, nos, os* (7a) are non-tonic SE-anaphors that adjoin to the flexive system (Reinhart & Reuland 1993 propose that SE-anaphors adjoin to the predicative head).
 - b. **tonic**: *mí, ti, sí*, are tonic SE-anaphors (7c), and their movement to adjoin the predicative head is prevented due to structural reasons (the presence of a preposition due to Case reasons).
- C) Reinhart & Reuland's (1993) Conditions A and B hold in Spanish.

The hypotheses are summarized in table (71) right below:

²⁰ Actually, these three hypotheses have already been assumed by several authors, see footnotes 21 and 22 below. The apportionment of this thesis is to build a model that subsumes the three hypotheses, and can account for all the facts related to reflexivization in Spanish, as well as the differences with other languages such as English and Dutch.

²¹ See Otreo (1999), Torrego (1995), among others.

²² See Rivero (2002) and Torrego (1995).

(71) Reflexivization in Spanish (revisited):

SPANISH (TRADITIONAL GRAMMAR)	SPANISH (HYPOTHESES)	
Clitics (<i>me,te,se,nos,os</i>)	non-tonic SE-anaphors (<i>me,te,se,nos,os</i>)	(IRVs)
Complex anaphor (<i>sí/mi/ti+mismo</i>) (<i>él/ella/yo+mismo</i>)	SELF-anaphors (<i>sí/mi/ti+mismo</i>) (<i>él/ella/yo+mismo</i>)	(nIRVs) (optionally IRVs)
Simple anaphor (<i>mí,tí,sí</i>)	tonic SE-anaphors (<i>mí,tí,sí</i>)	(nIRVs + PP)

In this chapter, I will analyze reflexive verbs in Spanish and provide a unified analysis of reflexivization in Spanish, English and Dutch based on Reinhart & Reuland's (1993) framework on reflexivization.

5.3.1. Inherent Reflexive Verbs, SE-anaphors and the Double Chain Condition

IRVs are a subset of agentive verbs (verbs that select an agent [+c+m] subject) and some examples can be found in (72). They undergo a reflexivization operation in the lexicon, defined in (66) and repeated below. This operation is a valence reduction operation (Reinhart 2002, Reinhart & Siloni 2005): the internal θ -cluster is eliminated as well as the verb's capacity of assigning accusative Case. Hence, the reduced verb will behave as an unergative verb in the syntax.

- (72) a. Arwen **se peinaba** ante la mirada de Aragorn
Arwen **se combed** before the sight of Aragorn
b. Frodo no **se lavó** durante su estancia en Mordor
Frodo no **se washed** during his stay in Mordor
c. Smeagol no podía **reconocerse** en el reflejo que el agua le devolvía
Smeagol no could **recognize-se** on the reflection that the water him_{clitic} gave

(66) Lexical reflexivization: (based on Reinhart & Siloni 2005)

- a. Transitive (basic) entry: $V_{acc} ([+c+m]_1, [-c-m]_2)$
b. Reflexivized entry: $R(V_{weak\ acc}) ([+c+m]+[-c-m])_1$
c. Syntactic realization: $DP_{[+c+m]+[-c-m]1} V$
d. Interpretation: $\exists e \lambda x [e=V \ \& \ [+c+m],e = x \ \& \ [-c-m],e = x]$

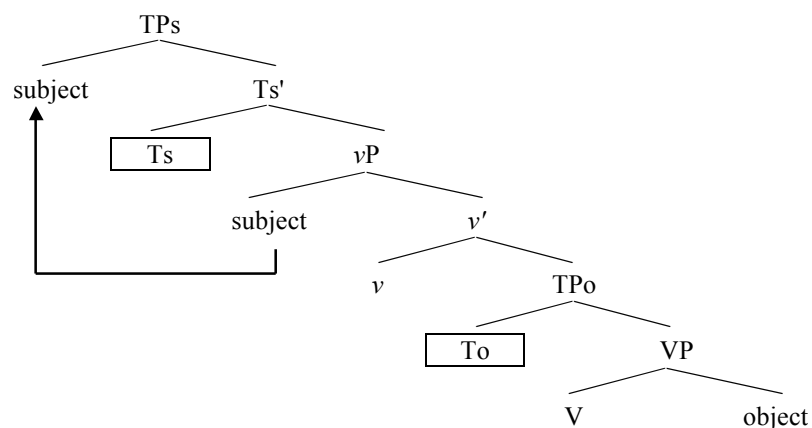
Since these verbs have undergone a lexical reflexivization operation, this is enough so as to mark the syntactic predicate as reflexive and fulfil R&R's B Condition²³. However, a SE-anaphor (*zich*) is required in Dutch (73a), as well as the clitic *se* (defined as a non-tonic or clitic SE-anaphor) in Spanish (73b).

- (73) a. Jan waste ***(zich)**
 Jan washed **zich**
 b. Juan ***(se)** lavó
 Juan **(se)** washed

The question arises as to why the presence of the SE-anaphor is required in Dutch and Spanish if the lexical reflexivization operation is enough to fulfil R&R's Condition B.

Recall that we have defined in chapter 4 the structure of the verbal predication following work by Pesetsky & Torrego (2004), and repeated here in (74):

- (74) Verbal predication structure: (Pesetsky & Torrego 2004:503)



There are two temporal heads: Ts licenses nominative Case and the subject, and To licenses accusative Case and the object.

My proposal is that Chomsky's (2001, 2006) view of phase heads and phase structure should be modified so as to capture the nature of these temporal heads. First, I will consider that the phase heads are those in (75b) and not those defended by Chomsky in (75a). By doing this, the phases are defined as full propositional units in the sense that all phase head (C and *v*) requires an argument (NP/DP) as well as a temporal head (Ts or To). Hence, the phase is a temporally specified unit with an argument (DP).

²³ In fact, languages like English use zero morphology with this type of verbs.

(75) a. Phase heads in Chomsky (2001, 2006):

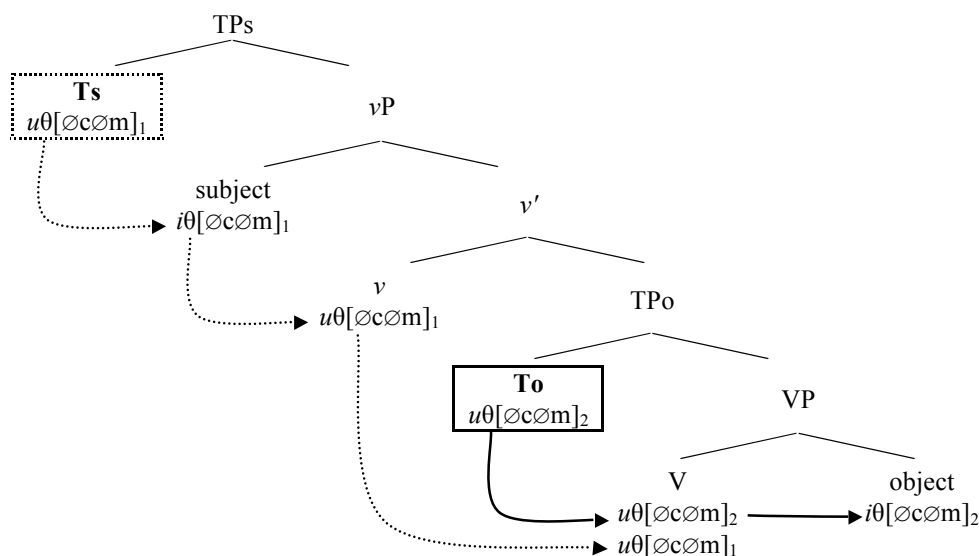
C - T y v - V

b. Phase heads (revisited):

C - Ts y v - To

The verb is the lexical head that introduces the values of the θ -features and the Tns-s and Tns-o features, besides the semantic and encyclopaedic contents of the event. Hence, it has to establish agree relations with its nominal arguments, as well as with the functional heads that define the phases. These relations are established by means of the *Agree* operation, which forms θ - and Tns-chains, schematized in (76):

(76) Relations between the temporal heads (Ts/To) and the verb:



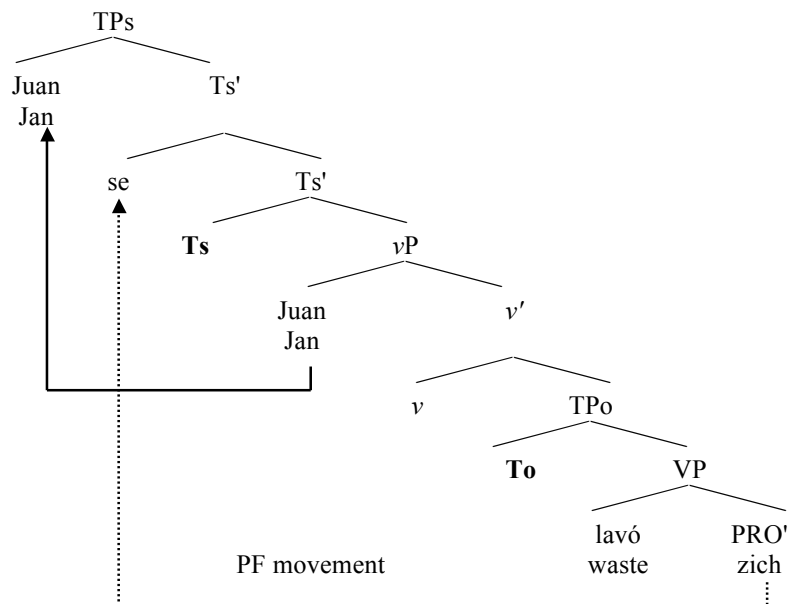
Note that both temporal heads Ts and To have uninterpretable and unvalued instances of theta-features. Furthermore, they are always included in the numeration of a sentence because both of them are necessary in order to form the temporal framework of the utterance (see chapter 2 section 2.3. above and Pesetsky & Torrego 2004).

When a sentence with an IRV is formed, just one nominal element is introduced in the numeration (the subject, which will be interpreted both as subject and object in the C-I system due to the lexical reflexivization operation). This causes a problem because in the internal phase²⁴ no nominal is introduced. Hence, the temporal head To cannot form part of a θ -chain with an interpretable instance of the θ -feature in order for the uninterpretable instance of the θ -features on To to be eliminated (note that the verb

²⁴ The terms *external phase* (CP) and *internal phase* (vP) are used for the sake of simplicity in the exposition.

does not have any instance of a θ -feature for the internal argument since it has undergone reflexivization). In order to prevent the derivation from crashing, the most simple SE-anaphor the UG has, is inserted: PRO'. This insertion is a last-resort mechanism. PRO' is not in the numeration, but it is inserted in the internal phase so as to form a θ -chain with To and hence, the uninterpretable instance of the θ -feature on To can be eliminated during the transference to the interfaces. Although PRO' ends up with its θ -features unvalued (unlike its φ -features), this causes no crash down because at the C-I interface all the uninterpretable instances of the θ -features have been deleted. I propose that an unvalued θ -feature can be tolerated by the C-I system (in a parallel way as unvalued φ -features are tolerated by the C-I system, see Holmberg 2005; and chapter 3 section 3.4).

(77) Juan se lavó / Jan waste zich.



How does the C-I system handle the unvalued θ -features? Is the SE-anaphor interpreted as a semantic argument of the verb? The answers to these two questions are related to each other and can be formalized in what I will call the *Double Chain Condition* (DCC) in (78). I have argued in chapter 3 (section 3.4) that an unvalued φ -feature is not problematic at the C-I interface: it is simply ignored, not interpreted at the C-I system. I will assume that the same happens with the unvalued θ -features: if they are unvalued, they are simply not interpreted at C-I. What happens then with an argument with unvalued θ -features? In the Government & Binding and Minimalist Program's literature, it has traditionally been argued that both structural Case assignment and thematic marking are essential for an argument to be interpreted as a semantic argument of the

verb. In other words, the intuition has been that an argument has to be both temporally (Case-marked) and thematically (θ -marked) integrated in the eventive structure. I propose to formalize this intuition in the Double Chain Condition in (78).

The DCC applies at the C-I system, not in C_{HL} . For an argument to be interpreted as a semantic argument of the verb, it needs to form θ - and tense-chains. What happens with the SE-anaphor inserted with IRVs for the sake of convergence at the interfaces? This anaphor forms a Tns-chain with To and fulfils (78b). However, it cannot form a θ -chain with the verb. The θ -chain formed is {To, PRO'}. Since the θ -features are interpretable in PRO', the uninterpretable instances on To can be deleted but no θ -chain is formed with the verb. PRO' ends up with its θ -feature unvalued and violates (78b), and hence, the DCC. As a result, PRO' is not interpreted as a semantic argument. However, it shares its φ -feature person with its antecedent. Therefore, its φ -feature person ends up valued and it is materialized as *me, te, se, nos, os*. The anaphor moves afterwards to the inflexion domain and hence its tonic nature.

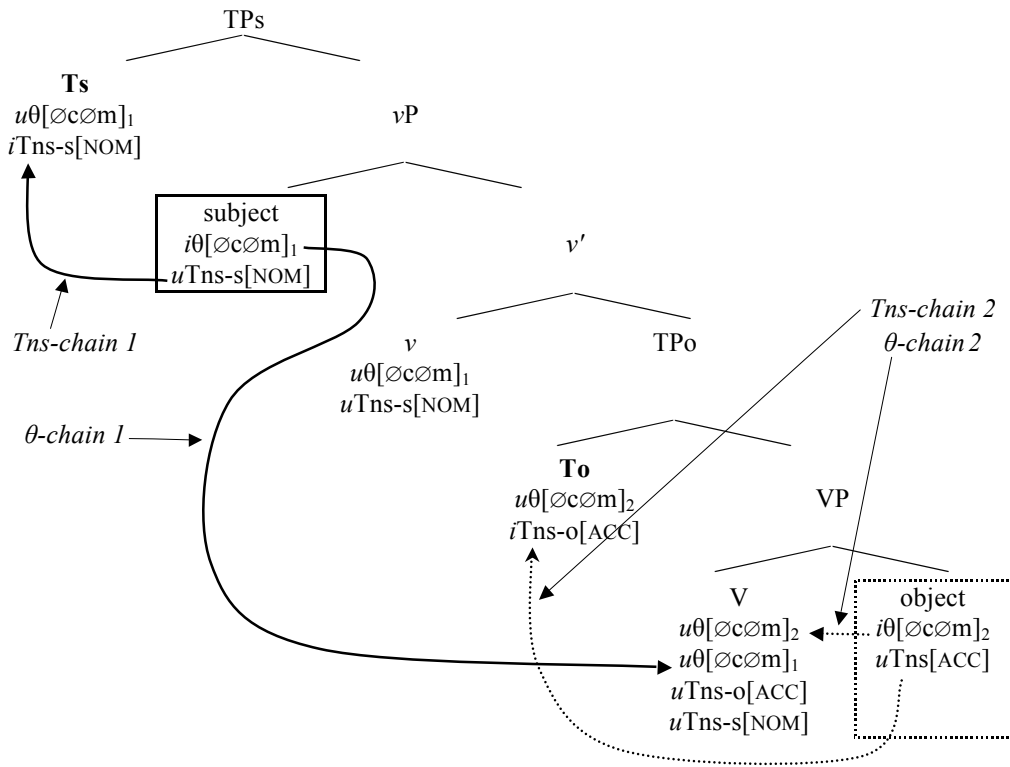
(78) The Double Chain Condition (DCC):

For a nominal item to be interpreted as a **semantic argument** of the verb at the C-I system, it has to form (by means of the *Agree* operation):

- a. one Tns-chain to share the Tns-s or Tns-o feature with the Ts or To heads
(or any other head that contains an interpretable Tns feature),
- b. and one θ -chain to share the θ -feature(s) with the verb.

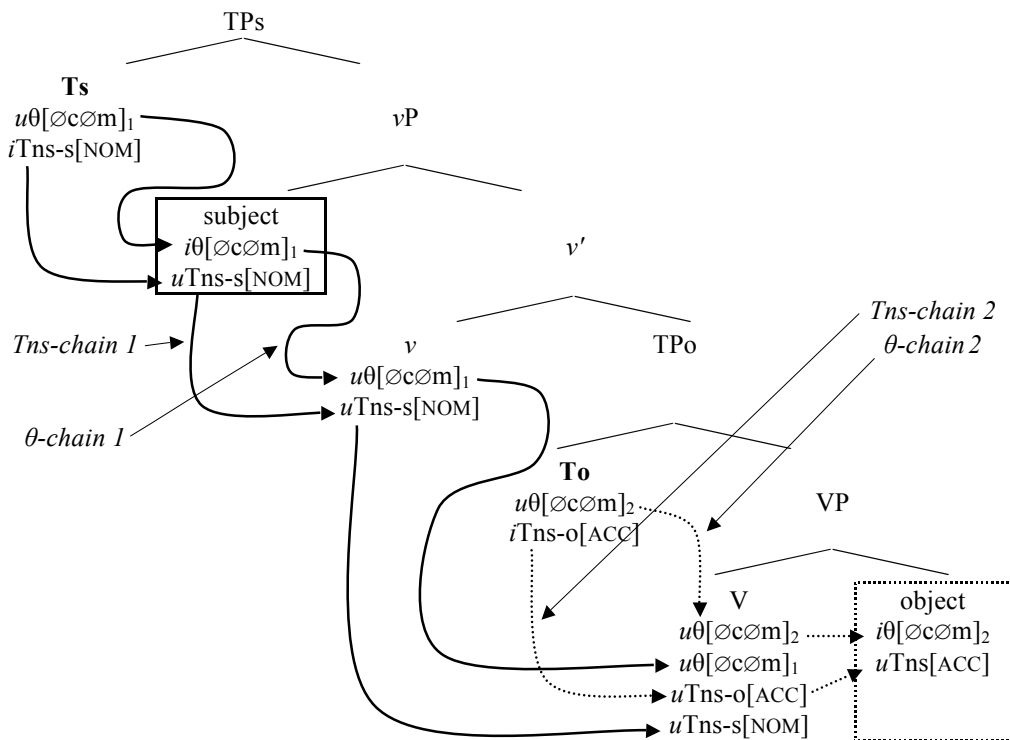
c.

(DCC: simplified agree-chains)



d.

(DCC: detailed agree-chains)



The semantic predicate has two objects since there are two variables bound by the same lambda operator. The syntactic predicate also has two arguments, the subject and the SE-anaphor. One of them is not interpreted as a semantic argument, though. However, the

verb behaves as unergative verb because it has an external argument. As for the internal argument, i.e. the anaphor, it does not form part of a θ -chain with the verb. In this sense, the anaphor is like an expletive similar to the English *it*.

Note finally, that the reflexivization operation needs to be reformulated. More concretely, the syntactic realization of the lexical entry has to specify the insertion of PRO'.

(79) Lexical reflexivization (revisited):

- a. Transitive (basic) entry: $V_{acc} ([+c+m]_1, [-c-m]_2)$
- b. Reflexivized entry: $R(V_{weak\ acc}) ([+c+m]+[-c-m])_1$
- c. **Syntactic realization:** $DP_{[+c+m]+[-c-m]1} V PRO'$
- d. Interpretation: $\exists e \lambda x [e=V \ \& \ [+c+m], e = x \ \& \ [-c-m], e = x]$

In sum, inherent reflexive verbs in Spanish are the outcome of a lexical reflexivization operation at the lexicon that takes a transitive verb entry as input and gives an unergative verb entry as output. This entry has an external θ -role, which is the result of bundling the external and the internal θ -roles of the transitive entry. Therefore, the insertion of PRO' is required in the syntax so as to delete the uninterpretable instance of the θ -feature on To. Since PRO' has its θ -feature unvalued, it does not fulfil the DCC and it is not interpreted as a participant of the event denoted by the verb.

5.3.2. Non Inherent Reflexive verbs and SELF-anaphors

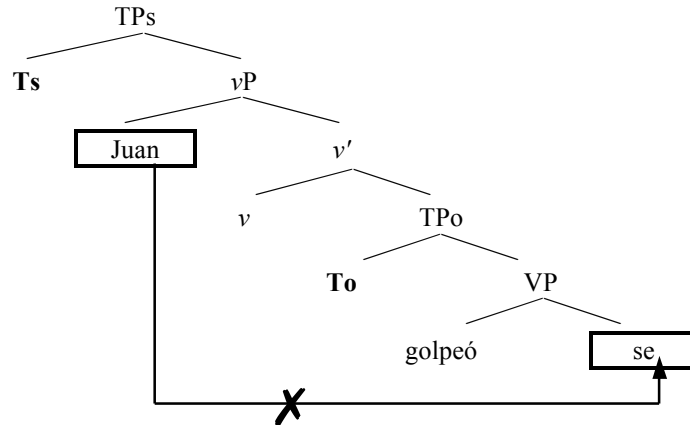
Virtually any verb in Spanish can reflexivize by inserting a SELF-anaphor, more concretely, causative verbs (those that select [+c] subjects). These verbs (nIRVs) enter the numeration with their valence unaltered, i.e. they do not undergo any kind of lexical operation that alters their valence.

The reflexive reading is derived from an A-binding process of the internal argument (the SELF-anaphor) by the external argument. I will call this A-binding *reflexive binding* because it gives a reflexive reading.

Since the verb does not undergo any reflexivization operation in the lexicon, the predicate needs to be marked as reflexive in the syntax so as not to violate R&R's Condition B. Therefore, a SELF-anaphor is needed. Recall that only this kind of anaphors (unlike SE-anaphors) can license reflexivization with nIRVs. If a SE-anaphor was inserted, this would be bound by the subject and the resulting chain would be translated to A-binding at the C-I interface (a desired result). However, the two

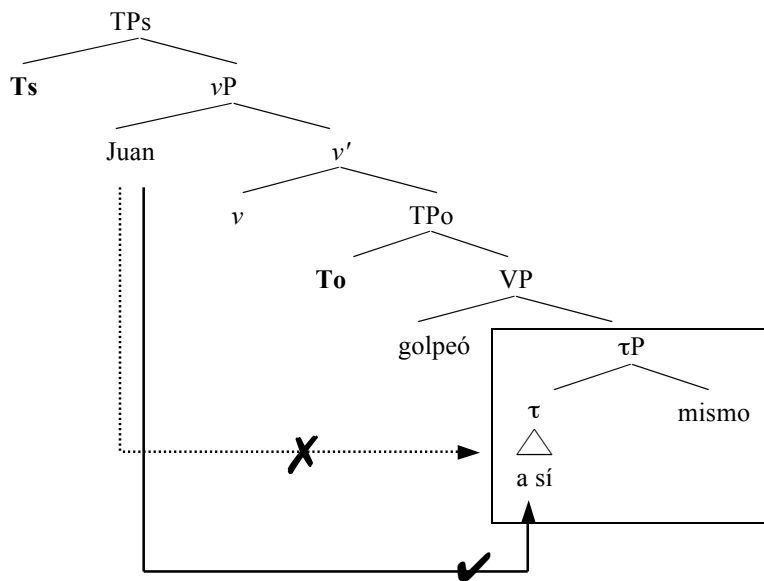
syntactic objects would be interpreted as just one semantic argument. This is due to the composition operation of agree-chains between the antecedent and the bound object, the SE-anaphor.

(80) *[Juan]_i [se]_i golpeó (reflexive reading, not inchoative)



On the other hand, the protector SELF element (MISMO in Spanish), prevents the formation of a chain between the antecedent and the bound constituent object τ in (81). Nevertheless, a chain is formed between the antecedent and the anaphor *sí*, which is within τ . Note however, that the chain is not formed between the antecedent and the bound object (τ). Therefore, there are two objects in the syntax that are translated to two distinguishable objects in the semantic.

(81) [Juan]_i [se]_i golpeó a [_i sí [_i mismo]]



The SELF-anaphor is duplicated with a clitic when it occupies an accusative or dative marked position, as well as other accusative or dative marked arguments (Torrego 1995).

- (82) Smeagol_i *(se) miró a **sí_i mismo** en el río
 Smeagol **se** saw to **sí self** on the river
"Smeagol saw himself on the river"

However, when the SELF-anaphor does not occupy an argumental accusative or dative marked position, it cannot be duplicated by the clitic, as with other arguments (Torrego 1995).

- (83) Frodo_i (*se) desconfía de **sí_i mismo** cuando lleva puesto el Anillo
 Frodo (se) distrust of **sí self** when bears on the Ring
"Frodo distrusts himself when he bears the Ring on."

Furthermore, in these cases, the SELF-anaphor can alternate with a tonic SE-anaphor. This is possible because in these cases, the SELF-anaphor does not mark the predicate as reflexive, and the introduction of a SE-anaphor does not violate R&R's B Condition.

- (84) Frodo_i vio ante **sí_i / sí_i mismo** al hombre que le robó el Anillo Único.
 Frodo saw before **si / sí self** to the man that him stole the One Ring
"Frodo saw the man that stole him the One Ring before himself."

- (85) a. Sam_i **se** critica a **él_i *(mismo)**
 Sam **se** criticizes to **him (self)**
 b. *[Sam]_i **se** critica a [él]_i
 Sam **se** criticizes to **him**
 c. [Sam]_i **se** critica a [_v él_i [_N mismo]]
 Sam **se** criticizes to **him** self

Finally, I claim (following Reuland 2001) that the element MISMO has a protective function with the SE-anaphor so as not to produce a mismatching with the valence of the verb, and thus to respect the Theta Criterion. It is a protector element with pronominals too. It allows them to be bound without violating Chomsky's Condition B (CBC). A pronominal can be in the determiner position of the τ phrase, and be bound by a local antecedent as in (85c). A pronominal without the protector element violates Condition B

(CBC) as well as R&R's Condition B. In other words, we see that a pronominal cannot mark a predicate as reflexive by itself.

5.3.3. Oblique *sí*

We have seen that SE-anaphors cannot be locally bound unless there is some reduction operation that reduces the arity of the verb. Otherwise, the Theta Criterion would be violated, as we can observe in (86).

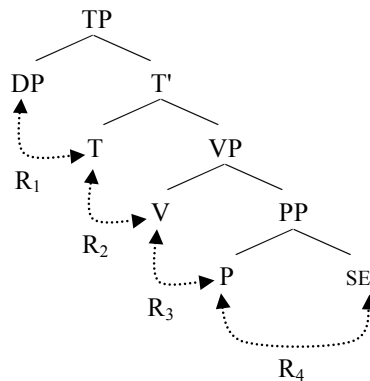
- (86) a. *Smeagol_i haat **zich_i**.
 Smeagol hates **zich**
 "Smeagol hates himself."
- b. *María_i **se_i** critica a **sí_i**.
 María **se** criticizes to **sí**
 "María criticizes herself."

In order to avoid the Theta Criterion violation the protector element SELF/MISMO is inserted so that the anaphor (the SE/SÍ/PRON) is A-bound not by means of chain formation but directly in the C-I system. The object τ is preserved, i.e. is not bound, and the Theta Criterion is respected as seen in (87).

- (87) a. Smeagol_i haat [_τ **zich_i** [_{N'} **zelf**]].
 Smeagol hates **zichzelf**
 "Smeagol hates himself."
- b. María_i **se_i** critica a [_τ **sí** [_{N'} **misma**]].
 María **se** criticizes to **sí misma**
 "María criticizes herself."

We have seen in chapter 3 that chain formation (and thus, binding in the syntax) within PP varies from language to language. In other words, some languages allow R₃ and/or R₄ in (88), whereas others do not.

(88)



If we assume that Spanish does not allow R_3 and/or R_4 in (88), the distribution of *sí* vs. *sí mismo* follows.

When *sí* is within a PP, there cannot be chain formation. The object is the PP but the binding is between the antecedent and *sí* within the PP. Since no chain composition is possible, then the binding takes place at C-I. The only restriction we can observe is that the antecedent has to c-command *sí*, as can be seen in (89), which also happens in NOC when it is done by means of A-binding at C-I (see chapter 3). The fact that *sí* needs to be c-commanded by its antecedent tells us that *sí* cannot get a value directly from discourse, i.e. it is [-R] as the rest of SE-anaphors (with the notable exception of PRO).

- (89) a. *[La madre [de Juan_i]] nunca habla mal de **sí_i**. Otero (1999:1447)
 [The mother [of Juan]] never speaks badly of **sí**
"The mother of Juan never speaks badly about him (him=Juan)."
- b. [La madre_i [de Juan]] nunca habla mal de **sí_i**. Otero (1999:1447)
 [The mother [of Juan]] never speaks badly of **sí**
"The mother of Juan never speaks badly about herself."

The fact that *sí* needs to be locally bound (it cannot be bound in domain 2, cf. chapter 3 section 3.3.2.3.) is due to the fact that Spanish allows only binding in domain for anaphors, unlike Dutch *zich* that allows binding in domain 2 as in (18) and (19) above and Italian that allows binding in domain 3 as in (20) above.

Note that dative and accusative *a* is not a preposition but a Case marker. Hence, chains can be formed and thus the protector element MISMO is needed as we saw in (86b) vs. (87b).

5.3.4. SELF-anaphors vs. SE-anaphors with Inherent Reflexive Verbs

The IRVs in Spanish can appear with or without SELF-anaphors. These anaphors are not necessary to fulfil R&R's Condition B since these verbs (can) undergo a lexical reflexivization operation, which marks the predicate as reflexive. Nevertheless, IRVs can appear with these anaphors. A verb like *lavar* (wash) can enter the numeration either reflexived or not, thus, it can enter the numeration with its valence either altered or unaltered. In the latter case, the transitive version of *lavar* can get a reflexive reading by A-binding of a SELF-anaphor.

- (90) a. Juan **se** lavó. *(lexical reflexivization version)*
 Juan **se** washed
 "*Juan washed*"
- b. Juan **se** lavó a **sí mismo**. *(transitive version)*
 Juan **se** washed to **sí self**
 "*Juan washed himself*"

The verb *lavar* is lexically reflexivized in (90a), whereas (90b) is the transitive version with a reflexive reading that comes from the A-binding of the SELF-anaphor.

It should be kept in mind that there are semantic differences between the lexical reflexivization and the reflexive binding versions. In other words, there are differences between SE-anaphors and SELF-anaphors. First, when the reflexivization operation applies, only one θ -cluster remains in the lexical entry of the verb. As a result of this, only one lambda operator will be present in the semantic representation, which will bind the two variables.

If there is no lexical reflexivization, the verb enters the numeration with its valence unaltered. Therefore, two lambda operators will be present in the semantic representation.

- (91) $x R f(x)$ *(Reuland 2001:481)*

The SELF/MISMO-element introduces an identity function defined in (91), by virtue of which the second semantic argument of R will be interpreted as a function of x . Therefore the variable y in the representation in (67c), repeated below, is interpreted as a function of the variable x (Reuland 2001). The consequence of this is that the two semantic objects are distinguishable although they will generally be interpreted as if

they were the same object, due to pragmatic reasons (see discussion on "Madame Tussaud" contexts below)

(67) Reflexive binding: (based on Doron & Rappaport-Hovav 2007)

- a. Lexical entry: $V_{acc} ([+c+m]_1, [-c-m]_2)$
- b. Syntactic realization: $DP_{[+c+m]} V DP_{[-c-m]}^{SELF-anaphor}$
- c. Interpretation: $\exists e \lambda x \lambda y [e=V \ \& \ [+c+m], e = x \ \& \ [-c-m], e = y \ \& \ y=f(x)]$

This difference, though small, can be perceived in ECM contexts in Dutch where both a SE-anaphor and a SELF-anaphor can appear. See the following example:

(92) “Madame Tussaud” context (89) in Reuland (2001:483)

Consider the following discourse in Dutch:

Marie is beroemd en liep bij Madame Tussaud’s binnen. Ze keek in een spiegel en

- a. ze zag zich in een griezelige hoek staan.
- b. ze zag zichzelf in een griezelige hoek staan.

Translation: Marie is famous and walked into Madame Tussaud's. She looked in a mirror and:

- a. she saw SE in a creepy corner stand. (i.e., she saw SE standing in a creepy corner)
- b. she saw herself in a creepy corner stand. (i.e., she saw herself standing in a creepy corner)

Favored interpretations:

- a. *zich* Marie: Marie saw herself
- b. *zichzelf* Marie's statue: Marie saw her statue

Reuland says the following regarding (92) (italics are mine):

"[In both contexts] *zich* gives an interpretation in which subject and object are identical. If *zichzelf* is chosen, subject and object are presented as distinguishable. In (89b) [(92b) here] the distinction is effected by interpreting *zichzelf* as a *representation of Marie* rather than as *Marie* itself. [...]

In both cases the sentence with *zichzelf* expresses a relation between an x and an $f(x)$ that bears a systematic resemblance to x , but can be distinguished from it. This implies that the structure in (88) [(92) here] is not just an

artifact of the analysis, but reflects a real property of *zichzelf* and its interpretation." (Reuland 2001:483)

This difference exists too in the Spanish sentences in (93), if the context allows it.

- (93) a. Juan **se** lavó. (lexical reflexivization version)
 Juan **se** washed
 "Juan washed"
- b. Juan **se** lavó a **sí mismo**. (transitive version)
 Juan **se** washed to **sí self**
 "Juan washed himself"

The sentence (93a) can only be interpreted as Juan washing himself, washing his body, whereas (93b) can also be interpreted, in a context similar to (56), as Juan washing an image of himself like a statue.

5.3.5. SELF-logophors

In section 5.2.1 we have seen examples of SELF-anaphors that are used as logophors in English (39b,c). The SELF-anaphors *myself* in (39b) and *himself* in (39c), repeated in (94) below, can be interpreted as logophors because they do not occupy an argumental position.

- (94) a. The queen invited [**me** / ***myself**] for tea. (R&R 1993:675)
 b. The queen invited both Max and [**myself** / **me**] for tea. (R&R 1993:675)
 c. Max_i said that the queen invited both Lucie and [**himself_i** / **him_i**] for tea. (R&R 1993:675)

The generalization that Reinhart & Reuland (1993) propose is that all the SELF-anaphors that do not mark a predicate as reflexive, are SELF-logophors. Therefore, a SELF-anaphor can always be used as logophor iff it is not in an argumental position. As logophors, they can have two functions (Reinhart & Reuland 1993:672; Reuland 2006c:2-4):

- a) **Perspective logophors**: like *himself* (94c), they express the viewpoint, the consciousness center, the perspective holder or the event orientation (Reuland 2006c:2-4).
- b) **Focus logophors**: their function is emphatic, like *myself* in (94b) (Reinhart & Reuland 1993:675).

It seems that in Spanish, the SELF-anaphors can only be focus SELF-logophors as in (95), but they can never be used to express event orientation (Teomiro 2008).

(95) Este libro lo escribí **yo / yo mismo**

This book *it*_{clitic} wrote **I / I self**

"I wrote this book myself"

The SE-anaphors are never focus logophors in Spanish, but they can express event orientation (as one of their interpretive possibilities). This issue is, nonetheless, beyond the scope of this thesis²⁵.

5.3.6. Recapitulation

We have seen that the reflexivization in Spanish can be accounted for by resorting to Reinhart & Reuland (1993) Conditions on reflexivization. On the one hand, Spanish, like Dutch (and English, as we will see in the next section) has SE-anaphors that can be tonic or non-tonic. These anaphors are needed with IRVs so as not to violate the Theta Criterion. On the other hand, SELF-anaphors are inserted with nIRVs in English, Dutch and Spanish, to license reflexive readings with such kind of verbs. The table in (96) presents a summary of this section.

(96) Reflexivization in English, Dutch and Spanish (revisited):

ENGLISH	DUTCH	SPANISH	
SE-anaphor (zero-morphology)	SE-anaphor (<i>zich</i>)	SE-anaphor (clitics: <i>me, se...</i>)	(IRVs)
SELF-anaphor (<i>himself</i>)	SELF-anaphor (<i>zichzelf</i>)	SELF-anaphor (<i>sí/mi/ti+mismo</i>) (<i>él/ella/yo+mismo</i>)	(nIRVs) (optionally IRVs)
		P+SE-anaphor (<i>mí, tí, sí</i>)	(nIRVs)

5.4. SE-anaphors at the syntax-phonology interface

As said before in section 5.3.1, both Spanish and Dutch insert SE-anaphors with IRVs to delete the uninterpretable instances of the θ -features on To.

²⁵ See Teomiro (2008: chapter 4) for a tentative analysis of the logophoric functions of the SE-anaphors in and their interpretation in impersonal constructions where they occupy the subject position in Spanish. However, I no longer maintain that analysis and I have reformulated the hypotheses on the status and function of the clitic *se* in Spanish impersonal and passive sentences in chapter 6.

- (97) a. Juan **se** lavó
 Juan **se** washed
 "Juan washed."
 b. Jan waste **zich**
 Jan washed **zich**
 "Jan washed."

English, on the contrary, uses zero-morphology with IRVs.

- (98) John washed \emptyset .

Note however, that the definition we have followed of lexical reflexivization requires the insertion of PRO' to delete uninterpretable θ -features on To (79c).

(79) Lexical reflexivization (revisited):

- a. Transitive (basic) entry: $V_{acc} ([+c+m]_1, [-c-m]_2)$
 b. Reflexivized entry: $R(V_{weak\ acc}) ([+c+m]+[-c-m])_1$
 c. **Syntactic realization:** $DP_{[+c+m]+[-c-m]I} V PRO'$
 d. Interpretation: $\exists e \lambda x [e=V \ \& \ [+c+m], e = x \ \& \ [-c-m], e = x]$

Therefore, I propose that the null SE-anaphor PRO' is present in English too, although it is not pronounced. So (98) would be (99):

- (99) John washed **PRO'**.

Whether *se*, *zich* or zero-morphology is used, is explained at the S-M interface. Each language chooses how to spell out the ϕ -features of its SE-anaphors. This is a process that takes place at the S-M interface, i.e. outside the narrow syntax or C_{HL} . For the time being, the spell-out rules in (100) allow us to explain the presence of *se* and *zich* in Spanish and Dutch, as well as the zero-morphology in English. In chapter 6, I will work these rules out so as to account for some contrasts concerning certain kinds of unaccusative verbs in Spanish, Dutch and English.

(100) SE-anaphors spell-out rules (to be revisited in chapter 6):

- a. Spanish and Dutch materialize the ϕ -features of PRO' iff they end up valued at the S-M interface.
 b. English never materializes the ϕ -features of PRO'.

5.5. Conclusions

In this chapter I have shown that Spanish makes use of both SELF-anaphors and SE-anaphors, just like Dutch. I have argued that the reflexive readings can be achieved by two routes. The first one is a reflexivization operation in the lexicon, which reduces one θ -cluster from the lexical entry of the verb and bundles it with the external θ -cluster. The other one is reflexive binding, which is a SELF-anaphor A-bound by an antecedent. When reflexivization applies in the lexicon, a SE-anaphor (PRO') must be inserted in the internal phase for the uninterpretable instance of the θ -features on To to be eliminated. By doing this, the derivation can converge at the interfaces. This SE-anaphor is spelled out as a clitic (in Spanish but not in Dutch) because its ϕ -feature person gets valued by the ϕ -features of the antecedent, and because the anaphor moves to the inflexion domain. When there is reflexive binding, the SELF/MISMO-element acts as a protector element so that no chain can be formed between the antecedent and the bound object (otherwise both syntactic elements would be translated to just one semantic element, and the derivation would crash due to a Theta Criterion violation). Within PPs, the SELF-element is not needed because the prepositions in Spanish do block chain formation. These two ways of getting a reflexive reading are not equivalent, and there exists semantic differences between them both. The SELF-anaphor can also have a logophoric function of focus in Spanish. This evidence points towards the conclusion that in Spanish the phenomenon of reflexivization takes place following the same principles as in other languages like English and Dutch. In other words, R&R's (1993) Conditions also hold in Spanish.

The table (101) shows the different pronouns along a continuum in which PRO' (the simplest and most versatile pronominal element the UG has) and the pronominals define the extremes. Also the SE-anaphors of the Spanish (*se*) and Dutch (*zich*), and the SELF-anaphors are included so that the reader has a unified view of the pronominal elements that have been analyzed in this work,

(101) PRO – Pronominals Continuum:

	SE-anaphors			SELF-anaphors	pronominals	
	PRO'	PRO	<i>se/sí</i>	<i>zich</i>	himself zichzelf	he/him/his
R	[-R]	[+R]	[-R]	[-R]	[-R]	[+R]
Φ-features	<i>i</i> φ person [-]	<i>i</i> φ person [-] <i>i</i> φ number [-]	<i>i</i> φ person [3 ^a]	<i>i</i> φ person [3 ^a]	<i>i</i> φ person [3 ^a] <i>i</i> φ gender [♂] <i>i</i> φ number [sing]	<i>i</i> φ person [3 ^a] <i>i</i> φ gender [♂] <i>i</i> φ number [sing]
Structural Case	<i>u</i> T [-]	<i>u</i> T [-]	<i>u</i> T [-]	<i>u</i> T [-]	<i>u</i> T [-]	<i>u</i> T [-]
Phonological content	no	no	yes	yes	yes	yes
Binding at C _{LH}	yes	yes	yes	yes	no	no
Reflexivizing function	no	no	no	no	yes	no

To conclude, PRO and the other SE-anaphors (PRO', *se*, *zich*) have in common that they can form agree-chains that can be translated to A-binding at the C-I interface. This cannot be done by means of SELF-anaphors (precisely to prevent this is why the protector element SELF/MISMO is used) or pronominals. Nevertheless, PRO differs from the other SE-anaphors in that, since it has a grammatical number feature (see footnote 19), it is [+R]. Spanish *se* and Dutch *zich* in reflexive contexts are [-R] since they are really an instance of PRO' with valued and spelled out φ-features (the latter not in the case of English). However, since they violate the DCC, they are not interpreted as semantic argument of the verb and hence, it is no longer crucial whether they are [+R] or [-R]. They could not get any referential value in the C-I system even if [+R].

Chapter 6

Pronominal verbs, variation and adjustment strategies at the interfaces

In this chapter I will address the issue of the presence of the SE-anaphors *se* in Spanish and the particle *zich* in Dutch with other kinds of verbs than those seen in chapter 5 (Inherent Reflexive Verbs).

In the first section, I will introduce the topic by giving some preliminary definitions, the basic data that will be accounted for as well as the theoretical questions to be answered and the hypotheses. In section 6.2 I will provide the theoretical background upon which the analysis presented in the rest of the chapter is based.

I will show in section 6.3 that there are verbs that, apart from inherent reflexive verbs, require the insertion of SE-anaphors both in Spanish and Dutch. I will account for why this is so, and why the scenarios where SE-anaphors are needed, are different in Spanish and Dutch. Finally, I will show that also English requires SE-anaphors with the same kinds of verbs. The difference is whether this anaphor is pronounced (Spanish and Dutch) or not (English).

In section 6.4 I will explain why other monadic verbs do not require nor allow the presence of the pronominal particle. In section 6.5 I will address the issue of the alternating pronominal verbs, i.e. those verbs that allow (but do not require) the presence of the pronominal particle. In section 6.6 I will deal with aspectual datives like those studied by Horn (2008), and finally, in section 6.7 I will explain why certain verbs are ambiguous between an inchoative and a reflexive reading.

6.1. Introduction

In this section, I will first define some basic concepts that will be used throughout this chapter. Afterwards, I will provide the reader with the basic data that will be accounted for in the analysis developed later on. I will specify the questions that will be answered by the analysis. The idea that I will put forward is that verbs that require *se* in Spanish have undergone the lexical operation *reduction* (based on Reinhart 2002), while verbs that disallow (or do not require) *se* have undergone a lexical operation based on Hale & Keyser's (2000) *conflation*.

6.1.1. Preliminaries

PRONOMINAL VERBS are those verbs for which the syntactic realization of their argumental structure requires the insertion¹ of a particle² that (seems to) lacks interpretation in the participant structure of the event.

- (1) a. Juan ***(se)** asustó. *(subject experiencer verb)*
 Juan **(se)** got scared
"Juan got scared."
- b. Ana **(se)** cayó de la silla.
 Ana **(se)** fell off from the chair
"Ana fell off the chair."
- c. Jan bedacht **zich**. *(subject experiencer verb)*
 Jan changed his mind **zich**
"Jan changed his mind."
- d. Jan scheert **zich** elke morgen. *(inherent reflexive verb)*
 Jan shaves **zich** every morning
"Jan shaves every morning."

PRONOMINAL ALTERNATION is the one in which the pronominal verbs that allow (but do not require or prevent) the presence of the pronominal particle participate. There are interpretive differences related with causation, mental involvement and aspect.

- (2) a. Juan cayó (5 metros) / (durante horas). *(activity)*
 Juan fell (5 metres) / (for hours)
"Juan fell (5 metres) / (for hours)."
- b. Juan **se** cayó (*5 metros) / (*durante horas) *(achievement)*
 Juan **se** fell (*5 metres) / (*for hours)
"Juan fell (5 metres) / (for hours)."
- (3) a. Jan bewoog **(zich)**.
 Jan moved **(zich)**
"Jan moved (by himself)."

¹ Along the syntactic derivation (I assume the pronominal particle is not in the numeration).

² The *pronominal particle*: *se* in Spanish and *zich* in Dutch.

b. De tafel bewoog (#**zich**).

The table moved (#**zich**)

"The table moved by itself."

CAUSATIVE ALTERNATION is the one in which the verbs whose subject may be the internal argument or the external [+c] argument (thus compatible with agents and natural causes) participate.

(4) a. El vaso *(**se**) rompió. (*causative alternation*)

The glass *(**se**) broke

"The glass broke."

b. La tormenta / la piedra / Juan rompió el vaso.

The storm / the stone / Juan broke the glass

"The storm / the stone / Juan broke the glass."

PSEUDO-CAUSATIVE / AGENTIVE ALTERNATION is the one in which participate the verbs whose subject may be the internal argument or an external argument that is not compatible with a natural cause (agentive alternation) (5).

(5) a. Los precios aumentaron este año. (*agentive alternation*)

The prices rose this year

"The prices rose this year."

b. *La crisis / los bancos aumentaron los precios este año.

*The crisis / the banks rose the prices this year

"The crisis / the banks have the prices risen this year."

6.1.2. Data

Both Spanish and Dutch make use of pronominal verbs. Spanish and Dutch Inherent Reflexive Verbs (IRVs) in (6) and Subject Experiencer Verbs (SEVs) in (7) are non-alternating pronominal verbs.

(6) Inherent Reflexive Verbs (IRVs):

a. Juan *(**se**) lavó.

Juan *(**se**) washed

"Juan washed."

b. Jan waste *(**zich**).

Jan washed *(**zich**)

"Jan washed."

(7) Subject Experiencer Verbs (SEVs):

a. Juan *(**se**) sorprendió.

Juan *(**se**) got surprised

"Juan got surprised."

b. Jan verbaasde *(**zich**).

Jan got surprised *(**zich**)

"Jan got surprised."

Cross-linguistic differences are observed in Theme Unaccusative Verbs (TUVs) as those in (8): they are non-alternating pronominal in Spanish (8a), whereas they are non-pronominal in Dutch (8b).

(8) Theme Unaccusative Verbs (TUVs):

a. El vaso *(**se**) rompió.

The glass *(**se**) broke

"The glass broke."

b. De vaas brak (***zich**).

The glass broke (***zich**)

"The glass broke."

In Spanish and Dutch, the externally caused unaccusative verbs like TUVs (8) seem to be pronominal, whereas the Internally caused unaccusative verbs ICUVs (9) seem to be non-pronominal.

(9) Internally Caused Unaccusative Verbs (ICUVs):

a. La rosa (??/***se**) floreció.

The rose (??/***se**) blossomed

"The rose blossomed."

b. De rose bloeide (***zich**).

The rose blossomed (***zich**)

"The rose blossomed."

Masullo (1999) argues that externally caused unaccusative verbs like theme unaccusative verbs (TUVs) in (8) require the clitic *se* because they are basically

transitive causative. The clitic incorporates in an aspectual head above the verb and absorbs accusative, and hence, the remaining nominall (the theme) has to raise to check nominative Case against Tense. The clitic *se* is a nominal head with very little semantic specification compatible either internal causation or, in the case of TUVs, inanimate external causation. Masullo further argues that internally caused unaccusative verbs (ICUVs) like (9) are inherent unaccusative verbs, i.e. they are listed in the lexicon as monadic verbs, and do not allow the presence of the clitic *se* nor have a transitive alternative. They are usually (though not exclusively) existential and presentational verbs.

Mendikoetxea (2000) argues that externally caused verbs like *romperse* ("break") in (8) (theme unaccusative verbs [TUVs] in our terms) are reflexive predicates, which are bi-eventive and dyadic, and with PRO as their external argument. Internally-caused verbs like *estallar* ("shatter") or *florecer* ("blossom") in (9) (internally caused unaccusative verbs [ICUVs] in our terms) are unaccusative predicates, which are monoeventive and monadic, and whose only argument is an internal argument.

Non-pronominal verbs like internally caused unaccusative verbs (ICUVs) in (9) do not normally participate in the causative alternation though some (10) seem to participate in the pseudo-causative alternation.

- (10) a. *La crisis / los bancos aumentaron los precios.
 *The crisis / the banks rose the prices
 "The crisis / the banks have the prices risen."
- b. Los precios aumentaron.
 The prices rose
 "The prices rose."

Note that, like internally caused unaccusative verbs (ICUVs), agentive unergative verbs (11), which license a cognate object, as well as consumption and creation verbs (12), which also license (but do not require) an object, are always non-pronominal (except the consumption verbs with ethical datives in (16)).

- (11) Agentive Unergative Verbs (AUVs):
- a. Juan corre (10 metros).
 Juan runs (10 metres)
 "Juan runs (10 metres)"

- b. Jan **heeft** gerend.
Jan **has** run
"Jan has run."
- c. Jan **is** naar de brug gerend.
Jan **is** to the bridge run
"Jan has run to the bridge."

(12) Consumption Verbs:

- a. Juan comió (una pizza).
Juan ate (one pizza)
"Juan ate (one pizza)."
- b. Jan heeft (een pizza) gegeten.
Jan has (one pizza) eaten
"Jan has eaten one pizza."

Spanish has a subset of alternating pronominal verbs that participate in the causative alternation (13), and another subset that do not (14), though there is some dialectal and diachronic variation.

- (13) a. ??/*La tormenta / ??/*el freno / el maquinista paró el tren.
??/*The storm / ??/*the brake / the engine driver stopped the train
"The storm / the brak / the engine driver stopped the train."
- b. El tren (**se**) paró.
The train (**se**) stopped
"The train stopped."
- (14) a. %Pedro / %/*el viento cayó el vaso.
%Pedro / %/*the wind fell the glass
"Pedro / the wind made the glass fall off"
- b. El vaso (**se**) cayó.
The glass (**se**) fell
"The glass fell off."

Dutch also has a (smaller) subset of alternating pronominal verbs (15), which is different from the Spanish one.

- (15) a. Jan / *de tafel bewoog (**zich**).
 Jan / *the table moved (**zich**)
"Jan / the table moved."
- b. Jan herstelde (**zich**).
 Jan recovered (**zich**)
"Jan recovered."

Finally, there are some triadic pronominal verbs (16) in Spanish that allow the presence of what I will call *aspectual se*, also in some dialects of English as in Horn's (2008) example in (16c) where *me* is quite similar to *se* in (16a,b)).

- (16) Aspectual se:
- a. Juan (**se**) comió la pizza.
 Juan (**se**) ate the pizza
"Juan ate (up) the pizza."
- b. Juan (**se**) leyó el libro.
 Juan (**se**) read the book
"Juan read the (whole) book."
- c. And now I've married **me** a pretty little wife. (Horn 2008:169)

6.1.3. Theoretical hypotheses

As I explained in chapter 2 section 2.3, I will pursue Chomsky's (2006) idea of a narrow syntax consisting of basic operations (*Agree* and *Internal/External-Merge*), and the interface between the syntax and the external systems, i.e. the Sensorio-Motor (S-M) system, the Conceptual-Intentional (C-I) system and the lexicon. As we saw in chapter 5, I will use Reinhart's (2002) conceptualization of the lexicon-syntax interface: the Theta-System is the system that interfaces between the system of concepts and the syntax, and it consist of lexical entry (the traditional "lexicon"), a set of marking procedures and a set of arity operations.

As for the cross-linguistic and intra-linguistic variation, I have put forward two basic hypotheses:

- i. The system of concepts, the narrow syntax and the semantic system are universal (Chomsky 2001, 2005, 2006; Reinhart 2002; Reinhart & Siloni 2005).
- ii. The cross-linguistic, diachronic and intra-linguistic variation should be accounted for at the different interfaces: lexicon-syntax and syntax-phonology.

6.1.4. Questions to answer

The questions that will be answered in this chapter are the following:

- A. Why do the pronominal verbs (inherent reflexive verbs [IRVs], subject experiencer verbs [SEVs], and theme unaccusative verbs [TUVs]) require the pronominal particle?
- B. Why do theme unaccusative verbs [TUVs] require the pronominal particle in Spanish unlike in Dutch?
- C. Following Reinhart (2002), inherent reflexive verbs (IRVs), subject experiencer verbs (SEVs), theme unaccusative verbs (TUVs) and internally caused unaccusative verbs (ICUVs) are all derived by means of an internal reduction operation (either reflexivization or decausativization) at the lexicon. Why then do inherent reflexive verbs (IRVs), subject experiencer verbs (SEVs) and theme unaccusative verbs (TUVs) require the pronominal particle whereas internally caused unaccusative verbs (ICUVs) prevent it?
- D. Why other monadic verbs like Agentive Unergative Verbs (AUVs) and consumption verbs (CVs) without object prevent the pronominal particle?
- E. How are the different pronominal alternations and their semantic differences (related to Jackendoff's (1987) *actor tier*) derived?
- F. How is the aspectual dative derived and how is it interpreted in the C-I system?

Questions (A-B) will be answered in section 6.3, questions (C-D) will be answered in section 6.4, question (E) will be answered in section 6.5, and finally, question (F) will be answered in section 6.6.

6.1.5. Working Hypotheses

The working hypotheses that will lead the analysis presented in this chapter are following:

- a. The pronominal particle is a null SE-anaphor inserted along the syntactic derivation in order to delete the uninterpretable instances of θ -features on one of the temporal heads (Ts or To).
- b. The cross-linguistic variation regarding the presence of the pronominal particle (e.g. theme unaccusative verbs [TUVs] in Dutch vs. Spanish) is due to different spell-out rules on the SE-anaphors at the syntax-phonology interface.

- c. Internally caused unaccusative verbs (ICUVs) do not require the pronominal particle because they have an interpretable θ -cluster (due to a *conflation* operation at the lexicon) that is able to delete the uninterpretable instances of θ -features on Ts or To.

6.2. Theoretical background: The lexicon-syntax interface and unaccusativity

I will follow Reinhart's (2002) Theta System as a conceptualization of the lexicon-syntax interface, which has been introduced in chapter 5, section 5.2.2.

We saw that the Theta System consists of lexical entries, marking procedures and a set of arity operations. Basically, there are two kinds of arity operations: first, those operations that reduce the verb arity, i.e. *reduction operations*. On the other hand, there are operations that augment the verb arity, they are *expansion operations*.

In chapter 5 section 5.2.2. we saw one reduction operation: the *reflexivization* operation, whose result is the reflexive entries of transitive agentive verbs.

- (17) Lexical reflexivization: (based on Reinhart & Siloni 2005)
- | | |
|-----------------------------------|---|
| a. (Basic) transitive entry: | $V_{acc} (u[+c+m]_1, u[-c-m]_2)$ |
| b. Reflexivization ³ : | $R_r(V) (u[(+c+m)+(-c-m)]_1)$ |
| c. Syntactic realization: | $DP_{[(+c+m)+(-c-m)]_1} \nu P$ |
| d. Interpretation: | $\exists e \lambda x [e=V \ \& \ [(+c+m)+(-c-m)], e = x]$ |

There is another reduction operation: *decausativization* (Reinhart 2002, Reinhart & Siloni 2005) that derives unaccusative entries from transitive causative verbs. This operation takes as input a transitive causative verb and deletes its external [+c] argument. Depending on whether the remaining argument is marked with 1 or with 2, this will merge as an external or an internal argument respectively.

Theme Unaccusative Verbs (TUVs) are derived by decausativization and are unaccusative because their remaining argument is a theme [-c-m] marked with index 2, so that it merges as a complement of the verb.

³ In Reinhart (2002) *reflexivization* deletes the accusative feature of the verb leaving a residue in some languages. This is reformulated by Reinhart & Siloni (2005) as *reflexivization* and *decausativization* deleting the thematic accusative case but leaving the structural accusative case (in some languages). I will not go deep into this question but restrict myself to stating that *reflexivization* and *decausativization* deletes the value of Tns-o (ACC) on the verb. Therefore, the remaining argument (both in reflexive and unaccusative verbs) gets a Tns-s (NOM) feature.

(18) Decausativization (theme unaccusative verbs): (based on Reinhart 2002)

- a. (Basic) transitive entry: $V_{acc} (u[+c]_1, u[-c-m]_2)$
 b. Decausativization: $R_d(V) (u[-c-m]_2)$
 c. Syntactic realization⁴: $DP_{[-c-m]_2} \nu P \overline{DP}_{[-c-m]_2}$
 c. Interpretation: $\exists e \lambda x [e=V \ \& \ [-c-m], e = x]$

On the other hand, Subject Experiencer Verbs (SEVs), which also are derived by decausativization from transitive causative verbs, are unergative because their remaining argument is an experiencer [-c+m], which is a mixed theta-cluster and hence it is not marked. Due to economy factors, it merges as an external argument.

(19) Decausativization (subject experiencer verbs): (based on Reinhart 2002)

- a. (Basic) transitive entry: $V_{acc} (u[+c]_1, u[-c+m])$
 b. Decausativization: $R_d(V) (u[-c+m])$
 c. Syntactic realization: $DP_{[-c+m]} \nu P$
 d. Interpretation: $\exists e \lambda x [e=V \ \& \ [-c+m], e = x]$

Finally, we will see one expansion operation: *causativization* (Reinhart 2002; Horvath & Sioni 2008a, 2008b). This operation takes as input a verbal entry and adds an agent [+c+m] argument. If the original entry has another agent [+c+m] argument, then it undergoes decausativization in the sense that the [+c] value of the /c feature becomes [-c], and the argument becomes an instrument [-c+m].

(20) Causativization (Reinhart 2002, Horvath & Sioni 2008)

- a. Basic entry (transitive): $V (u[+c+m]_1, u[-c-m]_2)$
 b. Causativization: $E_c(V)_{acc} (u[+c+m]_1, u[-c+m], u[-c-m]_2)$
 c. Syntactic realization: $DP_{[+c+m]_1} \nu P DP_{[-c-m]_2} DP_{[-c+m]}$
 d. Interpretation: $\exists e \lambda x \lambda y \lambda z [e=V \ \& \ [+c+m], e=x \ \& \ [-c+m], e=y \ \& \ [-c+m], e=z]$

6.3. Pronominal verbs

In this section I will analyse the pronominal verbs, i.e. those whose argumental structure requires the presence of the pronominal particle. These verbs can be divided in three

⁴ Mendikoetxea (2000) also argues that unaccusative verbs like *romper* (break) have a PRO as their external argument (at least) in Spanish (note however, that I assume that the anaphor that is inserted with these verbs is not PRO but PRO', see chapter 4 for more details on the differences between both anaphors).

types: *inherent reflexive verbs*, which we already saw in chapter 5, will be briefly review in section 6.3.1, *subject experiencer verbs* in section 6.3.2 and *theme unaccusative verbs* in section 6.3.3. I will provide a unified analysis based on processes at the lexicon-syntax interface. Finally, I will also account for the cross-linguistic differences found among Dutch, Spanish and English with an analysis based on processes at the syntax-phonology interface in section 6.3.4.

6.3.1. Inherent reflexive verbs

Previously we saw in chapter 5 that Inherent Reflexive Verbs (IRVs) are verbs for which the reflexive reading can be obtained without resorting to a SELF-anaphor (although the presence of such an anaphor is possible). This can be attested in Spanish (21a) and (22a), Dutch (23a) and (24a), and English (25a,c).

(21) a. Juan_[+c+m] *(se) lavó (a sí mismo).

Juan_[+c+m] *(se) washed (to sí self)

"Juan washed (himself)"

b. *El agua_[+c-m] / Juan_[+c+m] lavó a María.

*The water_[+c-m] / Juan_[+c+m] washed to María

*"*The water / Juan washed María"*

(22) a. Juan_[+c+m] *(se) peina.

Juan_[+c+m] *(se) combs

"Juan combs (his hair)"

b. *El viento_[+c] / Juan_[+c+m] peina a María.

*The wind_[+c] / Juan_[+c+m] combs to María

*"*The wind / Juan combs María"*

(23) a. Jan_[+c+m] waste *(zich) (zichzelf).

Jan_[+c+m] washed *(zich) (himself)

"Jan washed (himself)"

b. *Het water_[+c-m] / Jan_[+c+m] waste Marie.

*The water_[+c-m] / Jan_[+c+m] washed Marie

*"*The water / Jan washed Marie"*

(24) a. Jan_[+c+m] kamde *(zich).

Jan_[+c+m] combed *(zich)

"Jan combed (his hair)"

- b. *De wind_[+c] / Jan_[+c+m] kamde Marie.
 *The wind_[+c] / Jan_[+c+m] combed Marie
 "**The wind / Juan combs María"

- (25) a. John_[+c+m] washed (himself) this morning.
 b. *The water_[+c-m] / Mary_[+c+m] washed John this morning.
 c. John shaved (himself) this morning.
 d. *The knife / Mary shaved John this mornig.

In Reinhart & Reuland's (1993) terms, we can say that they are marked in the lexicon as reflexive, so that they fulfil R&R Condition B by themselves without the help of a SELF-anaphor.

These verbs participate in the agentive alternation, i.e. they have a transitive alternate that selects an agent [+c+m] subject and a theme [-c-m] subject. This can be seen in Spanish (21b) and (22b), Dutch (23b) and (24b), and English (25b,d).

We claimed in chapter 4 that, following Doron & Rappaport-Hovav (2007), Reinhart (2002) and Reinhart & Siloni (2005), these verbs were derived in the lexicon by means of a reflexivization operation, defined in (26). This operation takes as input a transitive agentive verbal entry and bundles the internal theta role with the external one, so that there is one compound theta role (agent and theme) that merges externally. The accusative feature of the verb is reduced.

- (26) Lexical reflexivization: *(based on Reinhart & Siloni 2005)*
- | | |
|-----------------------------------|---|
| a. (Basic) transitive entry: | $V_{acc} (u[+c+m]_1, u[-c-m]_2)$ |
| b. Reflexivization ⁵ : | $R_r(V) (u[(+c+m)+(-c-m)]_1)$ |
| c. Syntactic realization: | $DP_{[(+c+m)+(-c-m)]_1} \nu P SE$ |
| d. Interpretation: | $\exists e \lambda x [e=V \ \& \ [(+c+m)+(-c-m)], e = x]$ |

⁵ In Reinhart (2002) *reflexivization* deletes the accusative feature of the verb leaving a residue in some languages. This is reformulated by Reinhart & Siloni (2005) as *reflexivization* and *decausativization* deleting the thematic accusative case but leaving the structural accusative case (in some languages). I will not go deep into this question but restrict myself to stating that *reflexivization* and *decausativization* deletes the value of Tns-o (ACC) on the verb. Therefore, the remaining argument (both in reflexive and unaccusative verbs) gets a Tns-s (NOM) feature.

The problem was that if the lexical operation of reflexivization is enough to fulfil R&R's Condition B, what are the SE-anaphor *se* in Spanish and *zich* in Dutch for? Moreover, why these both languages require a SE-anaphor, whereas English does not?

Recall that we saw in chapter 5 that the temporal heads establish θ -chains with the verb. This implies that the head To has an uninterpretable instance of (a) θ -feature(s). Whenever a reduction operation (such as reflexivization) applies at the lexicon, only one nominal item is inserted at the numeration.

In order to eliminate this uninterpretable instance the most versatile and economical nominal item available in UG is inserted as last resort mechanism: the SE-anaphor PRO'. Hence, the uninterpretable instance of the θ -feature(s) on To forms an agree-chain with PRO', which has an interpretable (and unvalued) instance of the θ -feature(s).

Due to other agree relations, the ϕ -feature person of PRO' ends up valued: the interpretable and unvalued Tns and ϕ -features (person) of PRO' form a chain with the unvalued ϕ -feature person of the verb and the valued Tns feature of the verb. The derivation proceeds and once the subject is introduced in [Spec, vP], it forms a Tns-chain with *v* and with the verb. Thanks to this chain, the verb forms a chain with the subject in order to make deletable the rest of its ϕ -features (grammatical number and gender, those that PRO' lacks) and by virtue of this chain, all the ϕ -features of the verb and of PRO' get valued. Hence, PRO' is pronounced with the ϕ -feature person that concords with the ϕ -feature person of the subject.

PRO' gets its person ϕ -feature valued and it is spelled-out as a clitic (in Spanish though not in Dutch) since it moves to the head of the predication (Reinhart & Reuland 1993, this movement would be covert in Dutch).

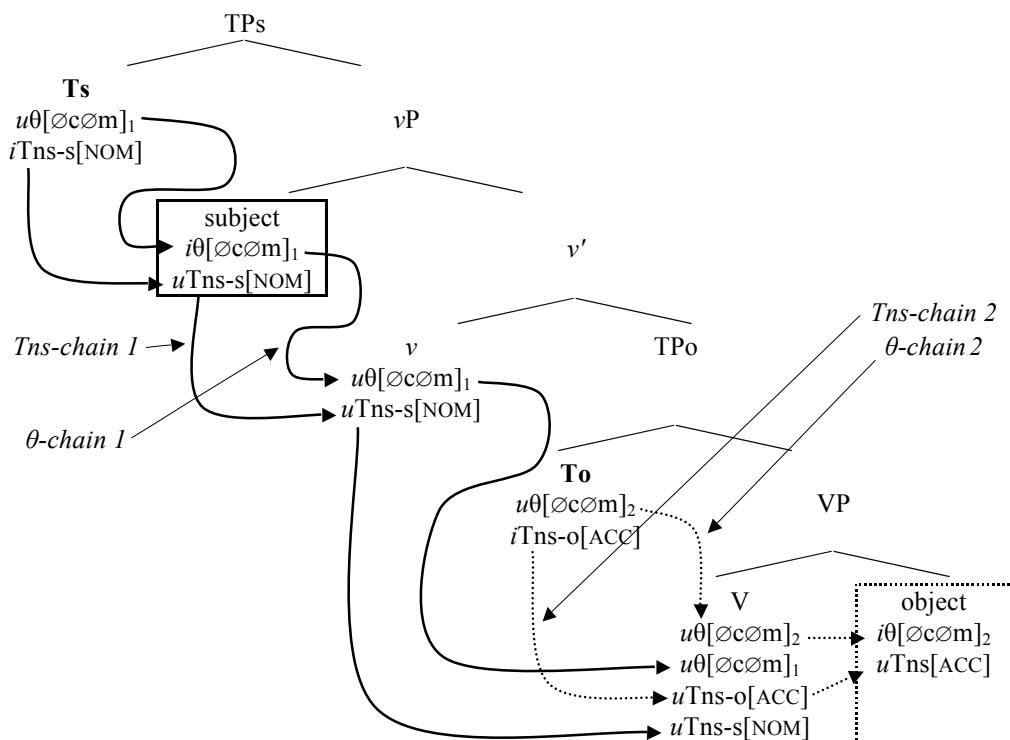
Nevertheless, the SE-anaphor PRO' is not interpreted at C-I as a semantic argument of the event denoted by the verb due to a DCC violation (27b). Recall that I claimed in chapter 3 that interpretability rather than valuation is crucial at the interfaces (see also Holmberg 2005, Epstein et al 2008).

(27) The Double Chain Condition (DCC):

For a nominal item to be interpreted as a **semantic argument** of the verb at the C-I system, it has to form (by means of the *Agree* operation):

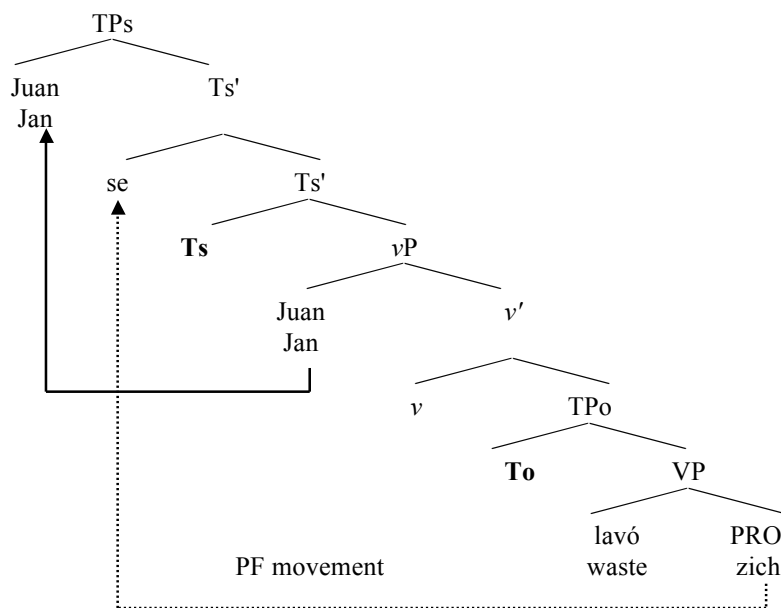
- a. one Tns-chain to share the Tns-s or Tns-o feature with the Ts or To heads
(or any other head that contains an interpretable Tns feature),
- b. and one θ -chain to share the θ -feature(s) with the verb.

c. (DCC: detailed agree-chains)



The derivation of a sentence with an IRVs such as (28) below, includes the merging of the null SE-anaphor PRO' in object position as a last resort mechanism. Since *Juan/Jan* establishes agree-relations with the verbal system, the person feature of PRO' ends up valued and spelled-out in Dutch (*zich*) and Spanish (*se*), though not in English (see section 6.3.4 below).

(28) Juan se lavó / Jan waste zich.



6.3.2. Decausativized experiencer verbs

Subject Experiencer Verbs (SEVs) are verbs that select an experiencer [-c+m] subject like (29a), (30a) and (33a) in Spanish, and (31a), (32a) and (34a) in Dutch. Some of them participate in the causative alternation, i.e. they have a transitive counterpart that select a [+c] subject (compatible with a cause, instrument or agent) and the [-c+m] argument is the internal one, as can be seen in (29b) and (30b) for Spanish, and (31b), (32b) for Dutch.

(29) a. Juan *(se) sorprendió.

Juan *(se) got surprised

"Juan got surprised"

b. El regalo / Juan sorprendió a María.

The present / Juan got surprised to María

"The present/Juan surprised María"

(30) a. Juan *(se) irritó.

Juan *(se) got angry

"Juan got angry"

b. El regalo / Juan irritó a María.

The present / Juan got angry to María

"The present/Juan got María angry"

- (31) a. Jan verbaasde *(**zich**).
Jan got surprised *(**zich**)
"Jan got surprised"
- b. Het cadautje / Jan verbaasde Marie.
The present / Jan surprised Marie
"The present / Jan surprised Marie"
- (32) a. Jan ergerde *(**zich**).
Jan got angry *(**zich**)
"Jan got angry"
- b. Het cadeautje / Jan ergeerde Marie.
The present / Jan got angry María
"The present/Juan got María angry"

A subset of subject experiencer verbs (SEVs) do not participate in the causative alternation, as can be seen in (33b) for Spanish and (34b) for Dutch.

- (33) a. Juan *(**se**) arrepintió.
Juan *(**se**) changed his mind
"John changed his mind"
- b. *El regalo / *Juan arrepintió a María
*The present / *Juan changed her mind to María
"The present / Juan made María change her mind"
- (34) a. Jan heeft *(**zich**) bedacht.
Jan has *(**zich**) changed his mind
"Jan changed his mind"
- b. *Het cadeautje / *Jan bedacht Marie.
*The present / *Jan changed her mind Marie
"The present / Jan made Marie change her mind"

According to Reinhart (2002), these subject experiencer verbs (SEVs) are derived entries from transitive causative verbs by a *decausativization* operation that applies at the lexicon, and it is defined in (35) for SEVs.

(35) Decausativization (subject experiencer verbs): (based on Reinhart 2002)

- | | |
|------------------------------|--|
| a. (Basic) transitive entry: | $V_{acc} (u[+c]_1, u[-c+m])$ |
| b. Decausativization: | $R_d(V) (u[-c+m])$ |
| c. Syntactic realization: | $DP_{[-c+m]} \nu P \mathbf{SE}$ |
| d. Interpretation: | $\exists e \lambda x [e=V \ \& \ [-c+m], e = x]$ |

Decausativization (35a) takes as input a verbal entry that selects a [+c] subject, thus, compatible with a cause, instrument or agent interpretation. The operation deletes the [+c] theta-feature altogether so that there is no trace of it in the semantic interpretation (35d). The remaining [-c+m] argument in (35b) is not marked since it is a mixed theta-cluster and it merges externally according to the merging instructions⁶. As it happens with the IRVs, there is only one argument in the numeration that merges externally and nothing within the νP can value (and make deletable) the uninterpretable theta-feature of To. Hence, PRO' is inserted in object position (35c) in order to delete the theta-feature of To, exactly as it happens with the IRVs.

The subset of subject experiencer verbs (SEVs) that do not participate in the causative alternation can be frozen decausativized entries (Reinhart 2002), i.e. only the decausativized entry of the verb has been grammaticalized. If so, we should be able to find causative entries in other languages (Reinhart 2002; Chierchia 1989).

Masullo (1999) argues that this subset of subject experiencer verbs that do not admit a causative alternation are inherent ergative verbs that generally express changes of position, disposition or physical or mental state, like *arrepentirse* ("change one's mind") and *enrojarse* ("reddening"). Masullo observes that these verbs can form adjectives and hence, he argues that they have two verbal layers, one expresses the cause and the other expresses the change of state. However, these verbs are compatible only with internal causation, unlike the theme unaccusative verbs (TUVs) above in (8), and therefore they only admit in the layer that expresses the cause the nominal head *se*, which is coindexed with the internal argument. Masullo says that, in a wide sense, this is a reflexive construction because both arguments are coindexed. However, only reflexive verbs like inherent reflexive verbs (IRVs) such as (6) admits expansion by means of insertion of a SELF-anaphor.

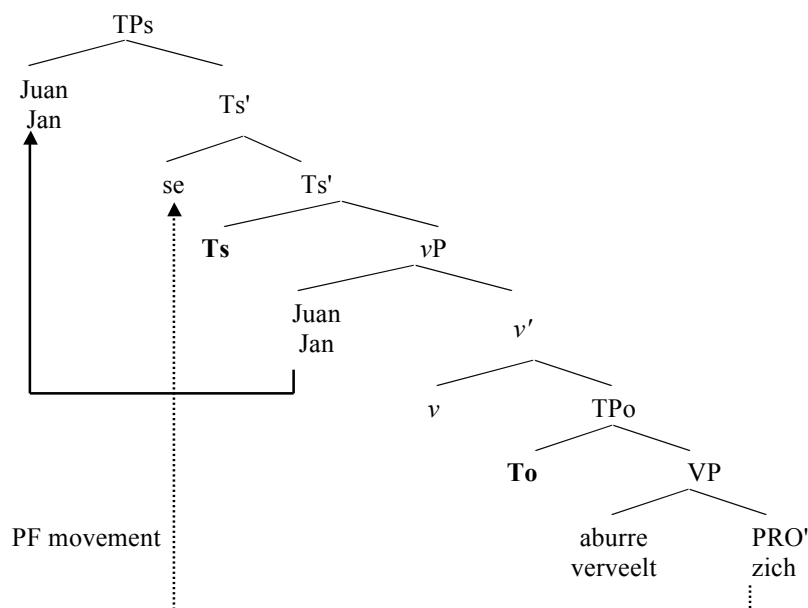
Another possibility is that their thematic grid is different, perhaps that in (36).

⁶ See chapter 4 for details on the marking procedures and the merging instructions of the Theta System.

(36) $V(u[+m]_1, u[-m]_2)$

A sentence like (37) below, requires the merging of PRO' as complement of the verb in order to delete the uninterpretable theta-features of To. Due to the agree relations of the subject [-c+m] with the verbal system, the person feature of PRO' is valued and ends up spelled out as *se* in Spanish and *zich* in Dutch.

(37) Juan **se** aburre / Jan verveelt **zich**.
 Juan **se** gets bored / Jan gets bored **zich**
"Juan/Jan gets bored."



6.3.3. Decausativized theme verbs

Theme Unaccusative Verbs (TUVs) select a theme [+c-m] argument that merges internally (within the vP as complement of the verb), and moves up to subject position ([Spec,Ts] in our framework) to check structural Case (nominative)⁷. Examples of TUV can be seen in (38a) and (41a) for Spanish, (39a) and (42a) for Dutch, and (40a) and (43a) for English.

(38) a. El vaso_[-c-m] *(**se**) rompió.
 El glass_[-c-m] *(**se**) broke
"The glass broke"

⁷ This is the standard analysis of unaccusative verbs in Generative Grammar since the work of Perlmutter (1978). See Levin & Rappaport-Hovav (1995) and Reinhart (2002), among many others.

- b. El viento_[+c-m] / Juan_[+c+m] rompió el vaso_[-c-m].
 The wind_[+c-m] / Juan_[+c+m] broke the glass_[-c-m]
"The wind / juan broke the glass"

- (39) a. De vaas_[-c-m] brak (***zich**).
 The glass_[-c-m] broke (***zich**)
"The glass broke"

- b. De wind_[+c-m] / Jan_[+c+m] brak de vaas_[-c-m].
 The wind_[+c-m] / Jan_[+c+m] broke the glass_[-c-m]
"The wind / juan broke the glass"

- (40) a. The glass_[-c-m] broke.

- b. The wind_[+c-m] / John_[+c+m] broke the glass_[-c-m]

Some TUV have a transitive causative alternate that selects a [+c] subject (compatible with cause, instrument and agent interpretation) and a theme [-c-m] object. This is what is known as causative alternation and can be seen in (38b) for Spanish, (39b) for Dutch, and (40b) for English.

There is a subset of TUVs that do not participate in the causative alternation: (41b) for Spanish, (42b) for Dutch and (43b) for English.

- (41) a. Juan_[-c-m] (**se**) ha marchado.
 Juan_[-c-m] (**se**) has left
"Juan has left"

- b. *La pena_[+c-m] / *Juan_[+c+m] ha marchado a María_[-c-m].
 *Sadness_[+c-m] / *Juan_[+c+m] has left to María_[-c-m]
"Sadness / Juan made María leave"

- (42) a. Jan_[-c-m] is (***zich**) weggegaan.
 Jan_[-c-m] is (***zich**) left
"Jan has left"

- b. *Het verdriet_[+c-m] / *Jan_[+c+m] heeft Marie_[-c-m] weggegaan.
 *Sadness_[+c-m] / *Jan_[+c+m] has Marie_[-c-m] left
"Sadness / Jan made Marie leave"

- (43) a. John_[-c-m] has left.

- b. *Sadness_[+c-m] / *John_[+c+m] left María_[-c-m] (made her leave).

- c. Sadness_[+c-m] / John_[+c+m] made María_[-c-m] leave.

According to Reinhart (2002), the TUVs are derived entries from transitive causative verbs by a *decausativization* operation that applies at the lexicon, and it is defined in (44) for TUVs.

(44) Decausativization (theme unaccusative verbs): (based on Reinhart 2002)

- | | |
|------------------------------|--|
| a. (Basic) transitive entry: | $V_{acc} (u[+c]_1, u[-c-m]_2)$ |
| b. Decausativization: | $R_d(V) (u[-c-m]_2)$ |
| c. Syntactic realization: | $SE DP_{[-c-m]_2} \nu P \overline{DP}_{[-e-m]_2}$ |
| c. Interpretation: | $\exists e \lambda x [e=V \ \& \ [-c-m], e = x]$ |

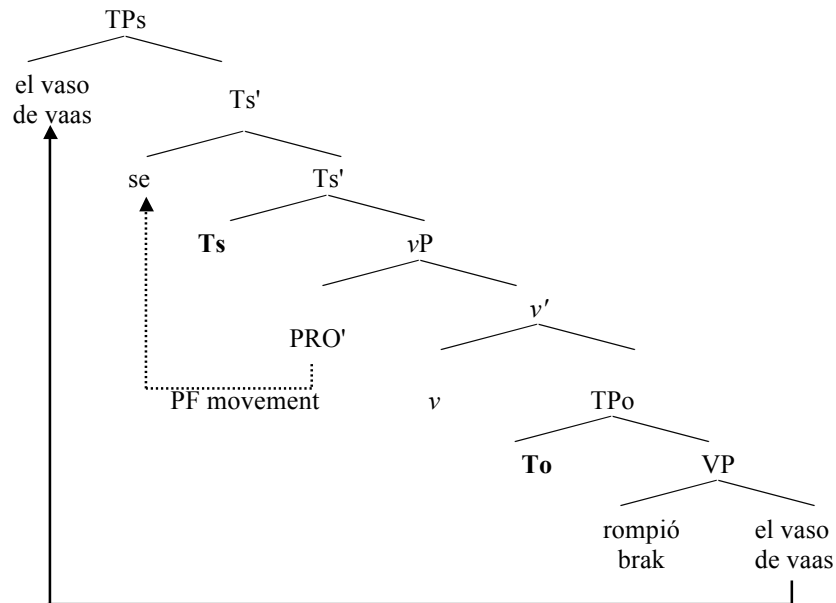
Decausativization (44a) takes as input a verbal entry that selects a [+c] subject, thus, compatible with a cause, instrument or agent interpretation. The operation deletes the [+c] theta-feature altogether so that there is no trace of it in the semantic interpretation (44d). The remaining [-c-m] argument in (44b) is marked with the index 2 since it is a [-] theta-cluster and it merges internally according to the merging instructions⁸. As it happens with the IRVs and SEVs, there is only one argument in the numeration although it merges internally, i.e. within the νP . However, nothing is merged in [Spec, νP] and hence, value (and make deletable) the uninterpretable theta-feature of Ts (rather than To, which appears with IRVs and SEVs). Hence, PRO' is inserted in subject position rather than object position (which happens with IRVs and SEVs), i.e. in [Spec, νP] (44c), in order to delete the theta-feature of Ts, in a similar way as it happens with the uninterpretable theta-features of To with IRVs and SEVs.

The subset of TUVs that do not participate in the causative alternation can be frozen decausativized entries (Reinhart 2002). As we said for SEVs, if this is on the right track, we should be able to find causative entries in other languages (Reinhart 2002; Chierchia 1989). Another possibility is that their thematic grid is different, probably some sort of two-place unaccusative verbs. However, I will not pursue this issue further here and leave this question open for future research.

A sentence like (45) below requires the merging of PRO' in [Spec, νP] in order to delete the uninterpretable theta-features of Ts. Due to the agree relations that it established with the verbal system, the person feature of PRO' is valued with the person feature of the argument *vaso/vaas* in complement position and ends up spelled out as *se* in Spanish but not as *zich* in Dutch, as it happens in English (see section 6.3.4 below).

⁸ See chapter 5 for details on the marking procedures and the merging instructions of the Theta System.

- (45) El vaso **se** rompió / De vaas brak (***zich**).
 The glass **se** broke / The glass broke (***zich**)
"The glass broke."



6.3.4. Cross-linguistic differences at the syntax-phonology interface

The issue at hand here is the cross-linguistic differences that are found in Spanish, English and Dutch. More concretely, IRVs require the presence of a SE-anaphor despite they are already marked as reflexive in the lexicon and fulfil R&R's Condition B by themselves. This SE-anaphor is inserted as it is with SEVs. However, TUVs require the insertion of the SE-anaphor PRO' too but only Spanish spells it out, in contrast with Dutch. Moreover, whereas both Spanish and Dutch spell out PRO' with IRVs and SEVs, English never spells PRO' out.

If we look closer at the structure of derivations with IRVs (46) and SEVs (47), we see that PRO' establishes a direct probe-goal relation with To. Hence, PRO' is marked with accusative Case. On the other, PRO' establishes a direct probe-goal relation with Ts rather than To with TUVs (48) (i.e. it is marked with nominative).

My claim is that this is the source of the cross-linguistic variation: depending on the relation of PRO', its person feature will be spelled out or not. Spanish spells out the person feature of PRO' whether it establishes a relation with To or with Ts. Dutch is more constrained and it spells out the person feature of PRO' iff it establishes a relation with To but not if it establishes a relation with Ts.

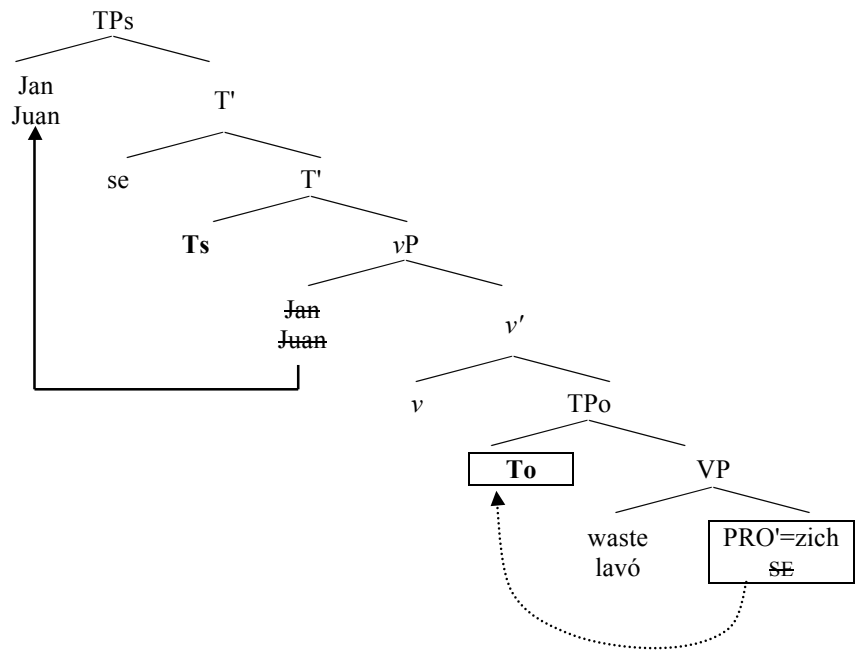
Finally, I claim that English is the more constrained of the three languages in the spelling out of the features of PRO': it uses zero-morphology whether PRO' establishes a relation with Ts or with To.

(46) IRVs:

Jan waste **zich** / Juan **se** lavó.

Jan washed **zich** / Juan **se** washed

"Jan/Juan washed"

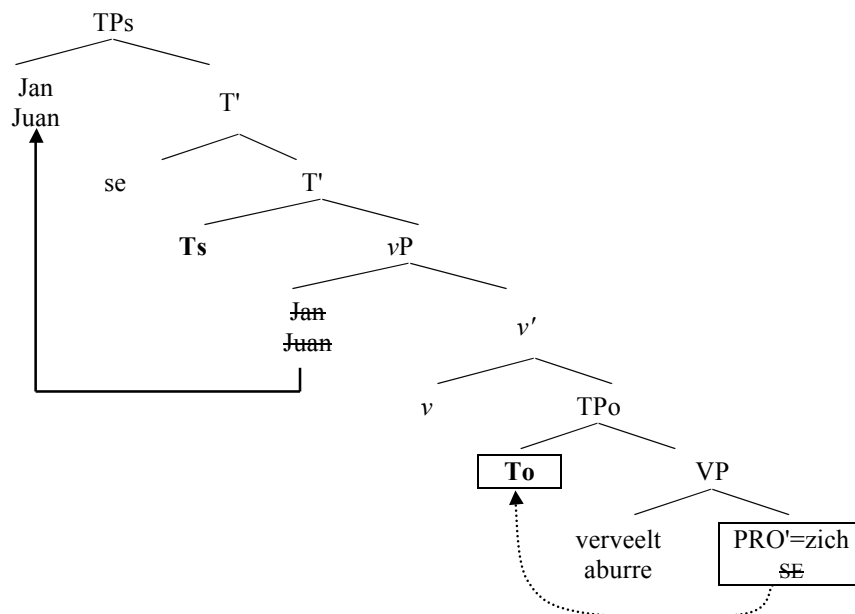


(47) SEVs:

Jan verveelt **zich** / Juan **se** aburre.

Jan gets bored **zich** / Juan **se** gets bored

"Jan/Juan gets bored"

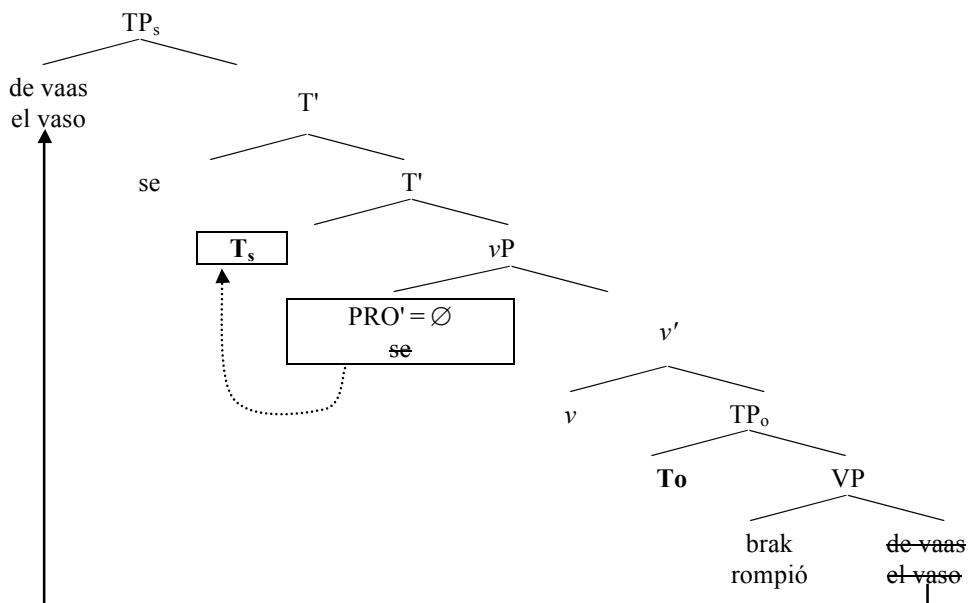


(48) TUVs:

De vaas brak (***zich**) / El vaso ***(se)** rompió.

The glass broke (***zich**) / The glass ***(se)** broke

"The glass broke"



We can summarize the claims above in the spell-rules in (49) right below:

(49) SE-anaphors spell-out rules:

- a. Spanish spells out the ϕ -feature (person) of PRO' iff it is valued.
- b. Dutch spells out the ϕ -feature (person) of PRO' iff it is valued and PRO' establishes a Tns-chain with To.
- c. English never spells out the ϕ -feature (person) of PRO'.

Finally, the table (50) summarizes the feature composition of the SE-features so far studied in chapters 5 and 6. This table will be revisited so as to accommodate PRO and other anaphors.

(50) Feature composition of SE-anaphors (revisited):

	PRO'	<i>se</i>	<i>zich</i>
ϕ -FEATURES	<i>i</i> ϕ person [unvalued]	<i>i</i> ϕ person [3 rd]	<i>i</i> ϕ person [3 rd]
θ -FEATURE(S)	<i>i</i> θ [unvalued]	<i>i</i> θ [unvalued]	<i>i</i> θ [unvalued]
TNS-FEATURE (CASE)	<i>u</i> Tns [unvalued]	<i>u</i> Tns [valued]	<i>u</i> Tns [valued]
PHONOLOGICAL CONTENT	no	yes	yes
R(EFERENCIAL INDEPENDENCE)	[-R]	[-R]	[-R]
TNS-CHAIN WITH	-	Ts/To	To

6.4. Non-pronominal verbs

Levin & Rappaport-Hovav (1995) and Reinhart (2002) assume that verbs like *blossom* are unaccusative verbs like TUVs such as *romper*, seen in section 6.3.3 above. However, there is one striking morphological difference between the former and the latter, at least in Spanish: whereas the former prevents the occurrence of the pronominal particle, TUVs require it. We will analyze in this section why this is so: whether it is a lexico-syntactic or a syntactic-phonological process (in other words, whether there is an unpronounced SE-anaphor). We will see verbs of the *floreecer*-type (blossom) in section 6.4.1. Verbs of the *augmentar*-type (augment), which seems to enter in the causative alternation, will be accounted for in section 6.4.2. Subsequently, existential and representational verbs will be handled in section 6.4.3. In section 6.4.4 I will analyse the agentive unergative verbs like *saltar* (jump) and *correr* (run) and what is the status of their cognate objects. The so-called *consumption verbs* such as *eat* will be analyzed in section 6.4.5. I will account for the status of their object and their aspectual properties.

Finally, I will explain the thematic grid of the verb *cambiar* (change) as a double marked for causation verb in section 6.4.6.

6.4.1. Internally caused unaccusative verbs: *florecer*-type verbs

Levin & Rappaport-Hovav (1995) argue that, although these verbs are unaccusative, they are not derived like TUVs but are basic entries. Reinhart (2002) argues that these verbs are derived entries since she argues that in Hebrew the transitive counterparts of almost all such verbs listed by Levin & Rappaport-Hovav can be found. Note that this kind of verbs, (51) for Spanish and (52) for Dutch, cannot enter in a real causative alternation: the less bad option is an instrument in Spanish (51b).

- (51) a. La rosa (***se**) floreció.
 The rose (***se**) blossomed
"The rose blossomed"
- b. ??El calor / *Juan floreció la rosa.
 ??The heat / *Juan blossomed the rose
"The heat/Juan blossomed the rose"
- (52) a. De roos bloeide (***zich**).
 The rose blossomed (***zich**)
"The rose blossomed"
- b. *De warmte / *Jan bloeide de roos.
 *The heat / *Jan blossomed the rose
"The heat/Juan blossomed the rose"

My claim is that these verbs are the result of the application of a new operation that I will call *conflation* (after Hale & Kayser 2000), (contra Reinhart 2002, who argues that they are frozen decausativized entries of externally caused verbs [+c])⁹.

The definition of this operation that I will make below is strongly based upon Hale & Kayser's (2000) *conflation* operation. I will assume that it takes place at the lexicon and that it is constrained by thematic reasons as *decausativization* is: whereas the latter is restricted to [+c] entries, *conflation* is restrained to [/-c-m] entries, i.e. it can apply to either [+c-m] or [-c-m] arguments.

⁹ If they were frozen entries, they would require the pronominal particle as it happens with non-alternating pronominal verbs (33) and (34), and (41) and (42).

Conflation applies to NPs. Recall from chapter 5 that NPs enter the numeration with an interpretable unvalued theta-cluster $i[\emptyset c \emptyset m]$. The application of conflation gives as result feature valuation as the operation *Agree* in the syntax. In other words, an NP with an interpretable and unvalued theta cluster $i[\emptyset c \emptyset m]$ in (53a) conflates with a verb that has an uninterpretable but valued theta cluster $u[/c-m]$ in (53b). The result of this operation is that a compound (NP+V) is formed in (53b). Within this compound, the NP has its interpretable theta-feature valued and, what is most important, the verb V has an uninterpretable valued theta-feature that has an interpretable occurrence. Hence, no argument needs to be introduced in the syntactic derivation to delete that uninterpretable theta cluster and the verb itself can delete the uninterpretable occurrences of theta-cluster either on Ts or To.

(53) [+c-m] conflation (I):

a. $NP_{i[\emptyset c \emptyset m]} \Rightarrow \text{Conflation} \Rightarrow V_{u[+c-m], u[-c-m]_2}$

b. $(NP+V)_{i[+c-m], u[-c-m]_2}$

The operation is formalized in (54) below. The input for the operation is a verbal entry with a $[/c-m]$ cluster (54a). The operation conflates a null NP with the verb, as specified in (53) and since the verb only has one uninterpretable occurrence of one theta-cluster (54b), only one argument enters into the syntactic derivation (54c) and no SE-anaphor is needed to delete the uninterpretable occurrences of the theta-features on either Ts or To.

(54) [+c-m] conflation (II):

a. Basic entry: $V_{acc} (u[+c-m], u[-c-m]_2)$

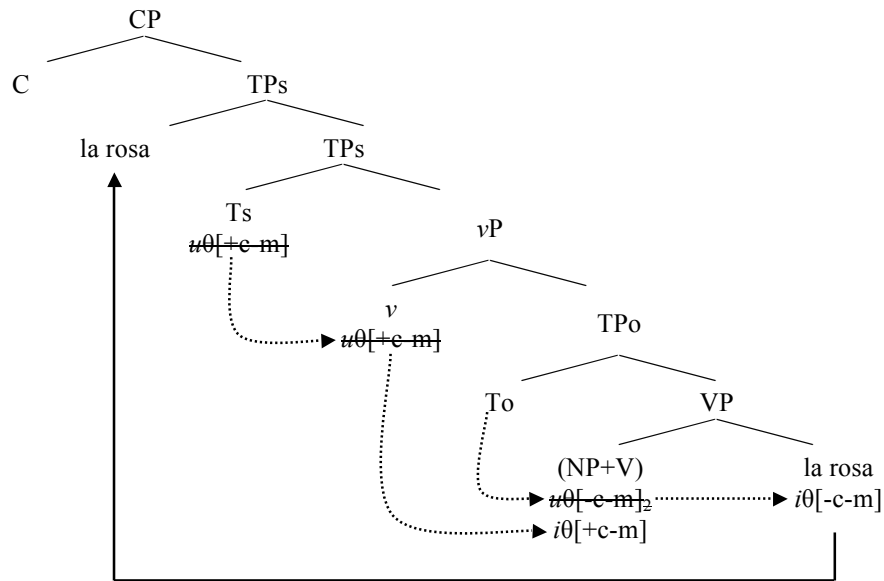
b. Incorporation: $V_{acc} (i[+c-m], u[-c-m]_2)$

c. Syntactic realization: $DP_{[-c-m]_2} \nu P \overline{DP}_{[-c-m]_2}$

d. Interpretation: $\exists e \lambda x \lambda y [e=V \ \& \ [+c-m], e=x \ \& \ [-c-m], e=y]$

In a sentence like (55) below, the pronominal particle is not required since the verb has an interpretable instance of θ , which can delete itself the uninterpretable occurrences of the θ -feature on Ts.

- (55) La rosa floreció.
 The rose blossomed
"The rose blossomed"



Note that the internal argument *la rosa* (the rose) shares its Tns-o feature with T_o, so in traditional terms, it checks structural accusative Case. However, it is morphologically realized as nominative due to the EPP feature of T_s. This can be seen in examples like (56), where *cinco duros* bears oblique case and nevertheless agrees (in grammatical number) with the inflected verb *bastan*.

- (56) Me bastan con cinco duros
 I_{dative} have_{plural-enough} with five duros (a kind of coin)
"I have enough with five "duros"."

6.4.2. Internally caused unaccusative verbs: *augmentar*-type verbs

These verbs are unaccusative but they (seem to) enter into the causative alternation as in (57) below: (57a) shows the unaccusative alternate of the verb *augmentar*, whereas (57b) shows the transitive version of such a verb. Note however, that the transitive version admits just agentive subjects [+c+m] and not subjects that can be interpreted as cause [+c] or instruments [+c-m].

- (57) a. Los precios (*se) aumentaron / disminuyeron.
 The prices (*se) rose / diminished
"The prices rose / diminished."

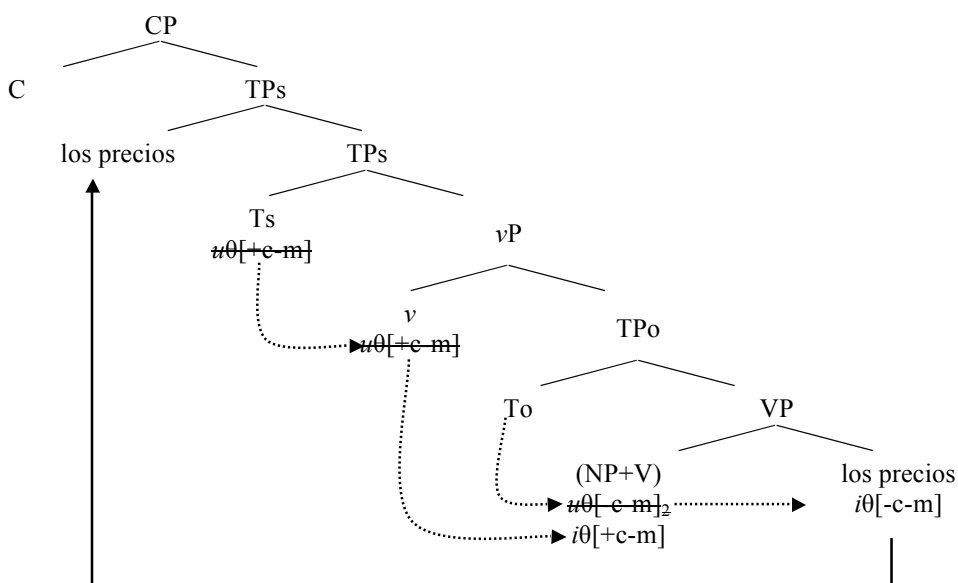
- b. *La crisis / los bancos aumentaron / disminuyeron los precios.
 *The crisis / the banks rose / diminished the prices
 "The crisis / the banks have the prices risen / diminished."

I claim that these verbs (*aumentar* and *disminuir*) are the result of the application of two operations at the lexicon: *causativization* (optional), and afterwards, *conflation* (again contra Reinhart 2002 for whom they are frozen decausativized entries of externally caused verbs [+c]).

The unaccusative version without agent (57a) is derived from *conflation*, formalized in (54) above. In a sentence like (57a), whose structure is represented in (58) below, PRO' is not necessary since the interpretable theta-features of the (NP+V) compound are able to delete the uninterpretable instances of the theta-features of Ts.

(58) Los precios aumentaron.

- The prices rose
 "The prices rose."



The transitive version with agent (57b) is derived from the application of *causativization*, which consists of adding an agent [+c+m] argument to a verbal entry at the lexicon, and thereafter *conflation* also at the lexicon. The formalization of these operations can be found in (59) below.

(59) Causativization and [+c-m] conflation:

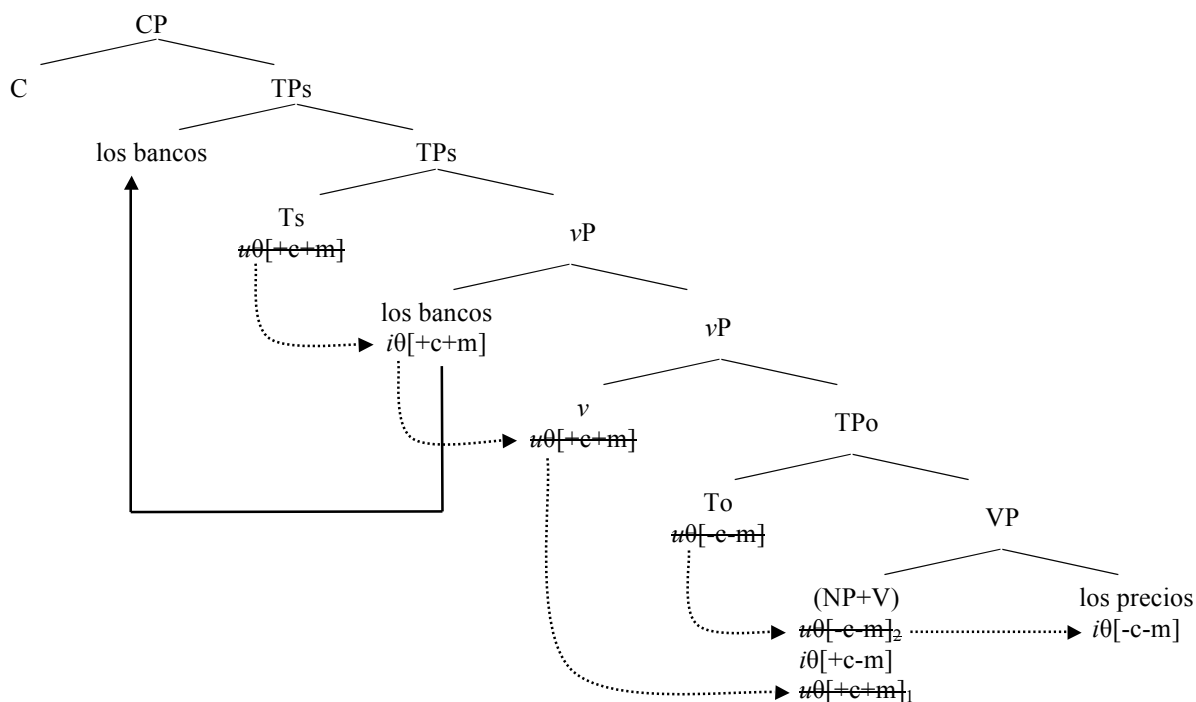
(based on Reinhart 2002
and Horvath & Siloni 2009)

- a. Basic entry: $V (u[+c-m], u[-c-m]_2)$
- b. Causativization: $E_c(V)_{acc} (u[+c+m]_1, u[+c-m], u[-c-m]_2)$
- b. Incorporation: $E_c(V)_{acc} (u[+c+m]_1, i[+c-m], u[-c-m]_2)$
- c. Syntactic realization: $DP_{[+c+m]1} vP DP_{[-c-m]2}$
- d. Interpretation: $\exists e\lambda x\lambda y\lambda z[e=V \ \&[+c-m],e=x \ \&[+c+m],e=y \ \&[-c-m],e=z]$

In a sentence like (57b), whose structure is represented in (60) below, the merging of PRO' is not necessary because the interpretable theta-features of the added agent [+c+m] delete the uninterpretable occurrences of theta-features on Ts. Besides the overt agent [+c+m], there is an implicit inanimate cause represented by the null NP with interpretable [+c-m] theta-features.

(60) Los bancos aumentaron los precios.

The banks rose the prices
"The banks have the prices rose."



6.4.3. Existential and presentational unaccusative verbs

My claim is that these verbs (61), (62) and (63) are also derived by the application of the operation *conflation* of two arguments [+c-m] and [-c] (locative) at the lexicon

(contra Reinhart 2002, who claims that they are frozen decausativized entries of externally caused verbs [+c]).

(61) Hay una niña (en el jardín).

There is a girl (in the garden)

"There's a girl (in the garden)"

(62) Er bestaan veel soorten planten (in de wereld).

There exist many sorts plants (in the world)

"There exist a lot of sorts of plants in the world"

(63) There's a girl (in the garden).

The locative argument [-c] is incorporated though it can be "duplicated". However, the overt locative argument is like a cognate object (see section 6.4.4) in the sense that the duplication would be a *locative restriction* and the overt locative cognate argument would be a *locative restrictor*.

The formalization of the conflation operation of the two arguments is given in (64) below. The result of the double conflation is a compound (NP+V+V_{loc}) with only one uninterpretable theta-cluster. Hence, these verbs require only one [-c-m] argument though they allow a locative argument, in our terms, a locative restrictor.

(64) [+c-m] and locative [-c] conflation (I):

a. $NP_{i[ec\circ m]} \Rightarrow \text{Conflation} \Rightarrow V_{u[+c-m], u[-c-m]_2, u[-c]_2}$

b. $(NP+V)_{i[+c-m], u[-c-m]_2, u[-c]_2} \Rightarrow \text{Conflation} \Rightarrow NP_{loc [ec\circ m]}$

c. $(NP+V+NP_{loc})_{i[+c-m], u[-c-m]_2, i[-c]_2}$

The operation that derives the existential and presentational verbs is formalized in (65) below. It takes as input a verbal entry with three arguments, two of which are [+c-m] and [-c]. These theta-clusters are conflated with two null NPs so that the compound (NP+V+NP_{loc}) require the merging of just one DP since there is only one uninterpretable theta-cluster.

(65) [+c-m] and locative [-c] conflation (II):

a. Basic entry: $V_{acc} (u[+c-m], u[-c-m]_2, u[-c]_2)$

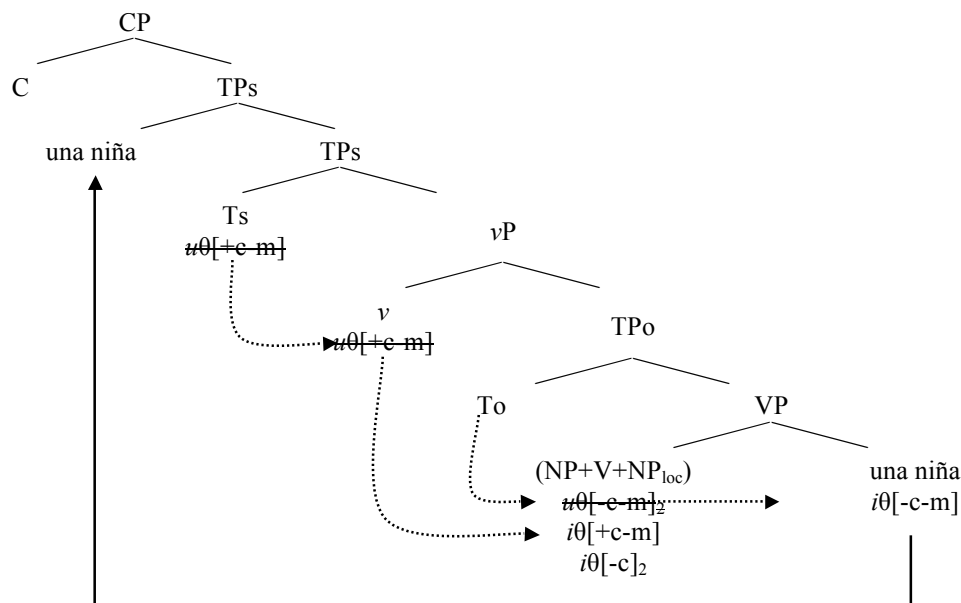
b. Incorporation: $V_{acc} (\underline{i[+c-m]}, u[-c-m]_2, \underline{i[-c]}_2)$

c. Syntactic realization: $DP_{[-c-m]_2} \nu P \overline{DP}_{[-c-m]_2}$

d. Interpretation: $\exists e \lambda x \lambda y \lambda z [e=V \ \& \ [+c-m], e=x \ \& \ [-c-m], e=y \ \& \ [-c], e=z]$

In a sentence like (66) below, only one argument is required. The uninterpretable instances of theta-features of Ts enter into an agree relation with the interpretable instances of the theta-features of the compound (NP+V+NP_{loc}). Hence, no SE-anaphor (PRO') is needed as a last resort mechanism.

- (66) a. Hay una niña.
 There is a girl
 "There's a girl"



6.4.4. Agentive unergative verbs

Agentive unergative verbs like (67) and (68) are monadic verbs, i.e. they select one argument, which has agentive properties (volitionality, control, instigation) and merges in subject position [Spec,vP] rather than in object position [V,DP] unlike unaccusative verbs.

These verbs optionally select an object, which is usually called *cognate object* because it is not compulsory and seems not to be part of the thematic grid (lexical entry) of the verb.

- (67) a. Juan (*se) corrió (la maratón / 10m).
 Juan (*se) ran (the marathon / 10m)
 "Juan ran"
 b. *la prisa / *Juan corrió a María.
 *Hurry / *Juan ran to María
 "Hurry/Juan made María run"

- (68) a. Jan rende (***zich**) (10m).
 Jan ran (***zich**) (10m)
"Jan ran (10m)"
- b. *Het haaste / *Jan rende Marie.
 *The hurry / *Jan ran Marie
"Hurry/Jan made Marie run"

Note that AUVs do not enter into the causative nor pseudo-causative alternation as seen in (67b) and (68b), i.e. they do not have a transitive counterpart that selects an external argument and the agentive argument [+c+m] merges within the vP.

My claim is that this kind of verbs is the result of the application of the operation *conflation* of the theme [-c-m] internal cluster at the lexicon, as formalized in (69) below.

- (69) [-c-m] conflation (I):
- a. $V_{u[+c+m],u[-c-m]_2} \Rightarrow \text{Conflation} \Rightarrow \text{NP}_{i[\theta_{c+m}]}$
- b. $(V+\text{NP})_{u[+c+m]_1,i[-c-m]_2}$

The operation takes as input a verbal entry with three arguments: one agent [+c+m], one theme [-c-m] and one optional subject matter [-m], as in (70a). The theme [-c-m] is conflated and the compound (NP+V) have only one uninterpretable theta-cluster. Therefore, only one argument is required to enter into the numeration. Since the remaining theta-cluster is an agent [+c+m] marked with an index 1, the argument that realizes it merges as an external argument.

Note that the pronominal particle is not required since the verb has an interpretable instance of θ , which can delete itself the uninterpretable feature on To.

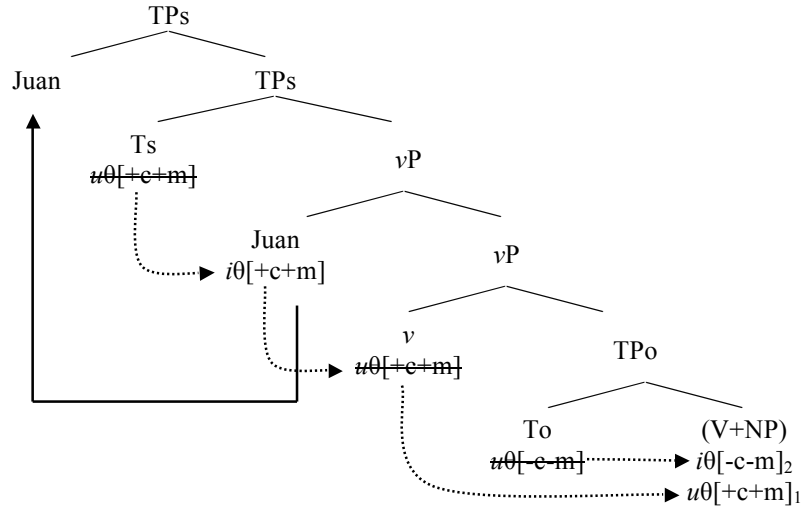
The cognate object is optional because it realizes the [-m] unitary theta-cluster rather than the theme [-c-m]. It shares its Tns-o feature with To, thus it bears structural and morphological accusative case. However, see in (70d) that in the semantics there is always an implicit [-c-m] argument, whether the cognate object [-m] is realized or not.

- (70) [-c-m] conflation (II):
- a. Basic entry: $V_{\text{acc}} (u[+c+m]_1, u[-c-m]_2 ([-m]_2))$
- b. Incorporation: $V_{\text{acc}} (u[+c+m]_1, i[-c-m]_2 ([-m]_2))$
- c. Syntactic realization: $\text{DP}_{[+c+m]_1} \text{ vP}$
- d. Interpretation: $\exists e \lambda x \lambda y (\lambda z) [e=V \ \& [+c+m], e=x \ \& [-c-m], e=y \ (\& [-m], e=z)]$

In a sentence like (71) below, only *Juan* needs to be merged along the syntactic derivation. The interpretable theta-cluster [-c-m] of the (V+NP) compound can value and make deletable the uninterpretable instances of theta-features on T_0 . Therefore, no SE-anaphor like PRO' is needed as last-resort mechanism.

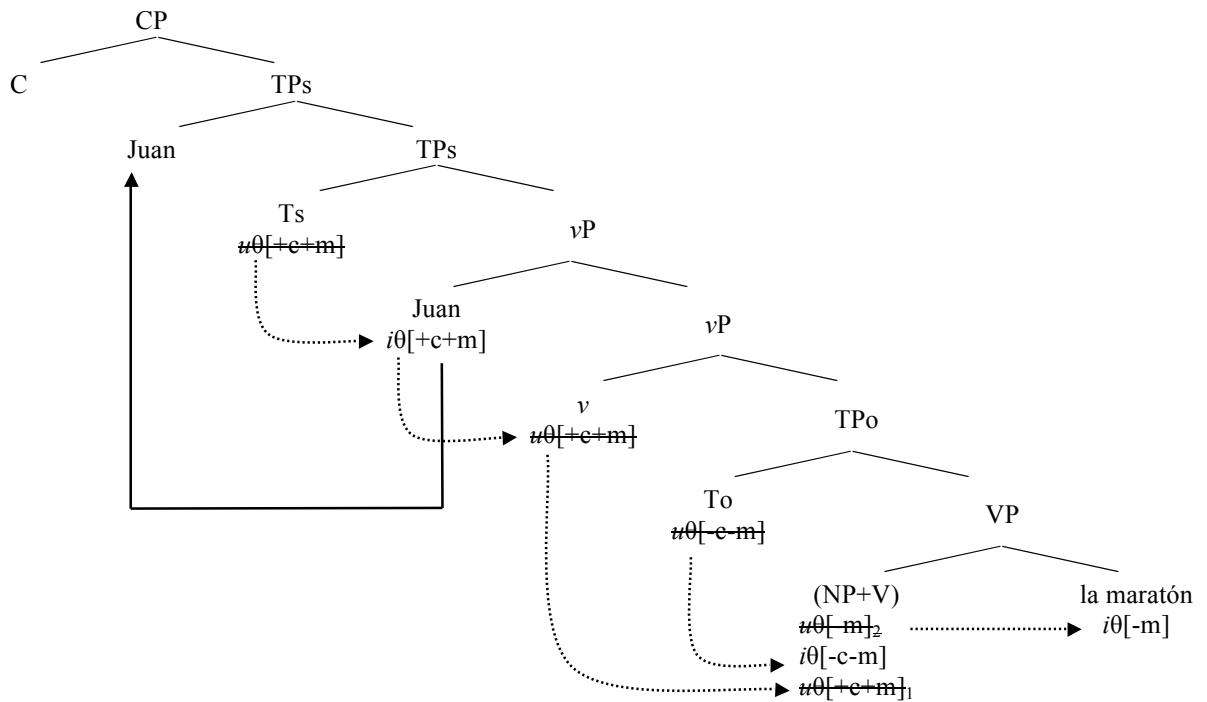
(71) Juan corrió.

Juan ran
 "Juan ran."



(72) Juan corrió la maratón.

Juan ran the marathon
 "Juan ran the marathon."



In a sentence like (72) above, *Juan* has interpretable instances of theta-features and then, it deletes the uninterpretable occurrences on Ts. The uninterpretable occurrences on To are deleted like in (71) above, i.e. by agreeing with the interpretable instances of the (V+NP) compound. The cognate object realizes the [-m] theta cluster, gets accusative Case and merges in object position.

6.4.5. Consumption verbs

Consumption verbs like (73a) and (74a) are assumed to be basically transitive, i.e. they select an agent [+c+m] subject and a theme [-c-m] object. However, and unlike other transitive verbs, the object can be left unpronounced.

- (73) a. Juan (***se**¹⁰) comió (la pizza).
 Juan (***se**) ate (the pizza)
 "Juan ate (the pizza)"
- b. *El hambre / *Juan comió a María.
 *The hungry / *Juan ate to María
 "Hungry/Juan made María eat"
- (74) a. Jan at (***zich**) (de pizza).
 Jan ate (***zich**) (the pizza)
 "Jan ate (the pizza)"
- b. *De honger / *Jan ate Marie.
 *The hungry / *Jan ate Marie
 "Hungry/Jan made Marie eat"

As seen in (73b) and (74b), this kind of verbs does not allow the causative nor pseudo-causative alternation, i.e. they do not have counterparts that select subjects and their agent [+c+m] merges in object position.

My claim is that these verbs are derived by the application of the operation *conflation* of the theme [-c-m] cluster at the lexicon, just like AUVs. The result is the (V+NP) compound with one interpretable theta-cluster and another uninterpretable one, as formalized in (75) below.

¹⁰ It would be grammatical if the pronominal particle is interpreted as an aspectual *se*, see section 6.6.

(75) [-c-m] conflation (I):

- a. $V_{u[+c+m],u[-c-m]_2} \Rightarrow Agree \Rightarrow NP_{i[-c-m]}$
- b. $(V+NP)_{u[+c+m]_1,i[-c-m]_2}$

The input of the operation is a verbal entry with three arguments, an agent [+c+m], a theme [-c-m] and an optional subject matter [-m], as in (76a). The theme [-c-m] argument is conflated and the (V+NP) compound has one interpretable theme [-c-m] cluster, that is implicitly understood as indicated in (76d), and two uninterpretable clusters: an agent [+c+m] and an optional subject matter [-m].

Only one argument is required by the argument structure of these verbs, the agent [+c+m]. The subject matter is optional, and it is the consumed object. The uninterpretable occurrence of theta-features on To are deleted by agreeing with the interpretable theta-cluster [-c-m] of the (V+NP) compound.

In other words, the overt object is not a theme [-c-m] (which is implicitly present since it is realized by the conflated null NP) but subject matter [-m]. In this sense, it is similar to a cognate object. Its interpretation is not as the object of the event denoted by verb but it is a restrictor of such event, hence its aspectual limitation. In this sense, it is similar to the locative [-c] in existential and presentational verbs in section 6.4.3: the cognate object is the restrictor of the event denoted by the verb.

(76) [-c-m] conflation (II):

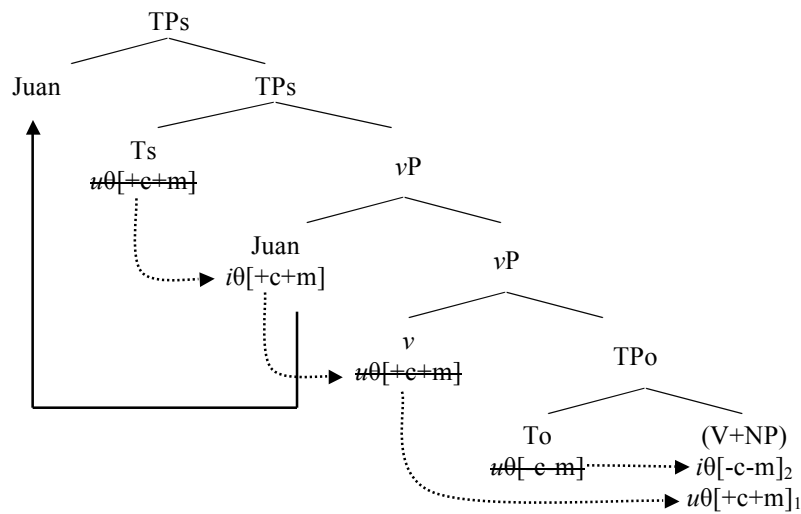
- a. Basic entry: $V_{acc} (u[+c+m]_1, u[-c-m]_2 ([-m]_2))$
- b. Incorporation: $V_{acc} (u[+c+m]_1, i[-c-m]_2 ([-m]_2))$
- c. Syntactic realization: $DP_{[+c+m]_1} vP$
- d. Interpretation: $\exists e \lambda x \lambda y (\lambda z) [e=V \ \& [+c+m], e=x \ \& [-c-m], e=y \ (\& [-m], e=z)]$

In a sentence like (77) below, only the argument *Juan* is required, and it deletes the uninterpretable instances of theta-features on Ts. The uninterpretable instances of theta-features on To are deleted by agreeing with the interpretable instance of the theta-cluster [-c-m] of the (V+NP) compound. Therefore, the interpretation of this sentence implies that there is one event of eating, of which the agent is *Juan* and what is eaten is not specified, i.e. it is somehow generic in the sense of "something is being eaten".

(77) Juan comió.

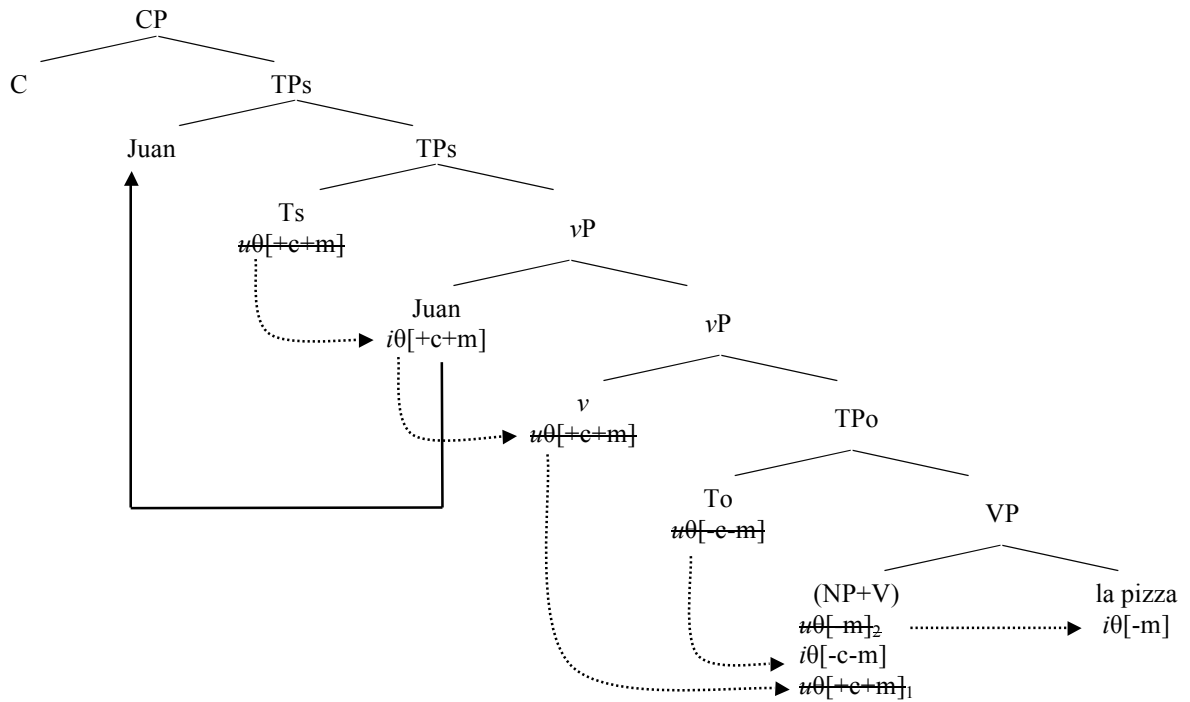
Juan ate

"Juan ate."



In a sentence like (78) below, something similar happens. Only the argument *Juan* is required, and it deletes the uninterpretable instances of theta-features on T_s. The uninterpretable instances of theta-features on T_o are deleted by agreeing with the interpretable instance of the theta-cluster [-c-m] of the (V+NP) compound. However, another argument is merged as internal argument (as complement of the verb) and realizes the [-m] theta-cluster. Therefore, this argument *pizza* restricts the event of eating. In other words, the semantics of this sentence implies that there is an event of eating, of which the agent is *Juan* and this event of eating is restricted by the object *pizza* since what is being eaten, the argument [-c-m], cannot be interpreted as generic but is restricted by the [-m] argument *pizza*.

- (78) Juan comió la pizza.
 Juan ate the pizza
 "Juan ate the pizza."



6.4.6. Double caused verbs: *cambiar*

I claim that verbs like *cambiar* (change) are specified with two causes in their verbal entry, which is specified in (79) below:

- (79) $V(cambiar)_{acc} u[+c]_1 i[+c-m] u[-c-m]_2$

The argument [+c-m] is conflated and the [+c] may be realized or not, due to the presence of the [+c-m] argument. The pronominal particle is not required to delete the θ-features on either T_s or T_o due to the interpretable theta-cluster of the (NP+V) compound.

6.5. Alternating pronominal verbs

Alternating pronominal verbs allow (though not require) the realization of the pronominal particle. There are subtle semantic and aspectual differences depending on whether the pronominal particle is realized or not. Here, I will deal with semantic differences related to the causal status of the participant the argument realizes (those related with the /c or /m features), leaving aside the aspectual differences (see De Miguel & Fernández Lagunilla 2000).

Examples of different alternating pronominal verbs are shown in (80)-(86), which will be analyzed in this section.

- (80) a. Juan **se** cayó (*5 metros). *(theme-subject matter verb)*
Juan **se** fell (*5 metros)
"Juan fell (5 metres)"
- b. Juan cayó (5 metros).
Juan fell (5 metres)
"Juan fell (5 metres)"
- (81) Juan (**se**) murió. *(theme-subject matter verb)*
Juan (**se**) died
"Juan died"
- (82) El tren (**se**) paró. *(theme-subject matter verb)*
The train (**se**) stopped
"The train stopped"
- (83) a. Ana (**se**) ha envejecido mucho. *(double caused [+c][-c] verb)*
Ana (**se**) has got older a lot
"Ana has got older"
- b. La mesa (***se**) ha envejecido mucho
The table (***se**) has got older a lot
"The table has got older"
- (84) a. Juan (**se**) ha mejorado / empeorado mucho. *(double caused [+c][-c] verb)*
Juan (**se**) has got better / worse a lot
"John has got better/worse a lot"
- b. La situación (***se**) ha mejorado / empeorado mucho.
The situation (***se**) has got better / worse a lot
"The situation has got better/worse a lot"
- (85) a. Jan bewoog (**zich**). *(externally caused [+c][-c] verb)*
Jan moved (**zich**)
"Jan moved"
- b. De tafel bewoog (**#zich**).
The table moved (**#zich**)
"The table moved"

- (86) Jan herstellt (**zich**). *(externally caused [+c][-c] verb)*
 Jan recovers (**zich**)
"Jan recovers"

6.5.1. Verbs with animacy restrictions: [+c][-c] verbs

The [+c][-c] verbs are pronominal or alternating pronominal verbs that impose an animacy restriction on their internal argument. They have an optional [-c] argument that fulfils the DCC (27), i.e. it is interpreted as an argument at the C-I system. Their basic entry is formalized in (87) below.

- (87) $V_{acc} [+c]_1[-c-m]_2 ([-c])_2$

Two kinds of [+c][-c] verbs can be distinguished: first, *mejorar(se)/(zich)bewegen*-type as (84) and (85), and second *envejecer(se)*-type as (83).

6.5.1.1. Mejorar(se) / (zich) bewegen-type verbs

They are pronominal in Spanish and alternating pronominal in Dutch: *moverse* vs. *(zich) bewegen* in (85) and *mejorarse* in (84) (with *se* requires animate arguments).

They allow causative alternation:

- (88) Juan / la crisis ha mejorado / empeorado la situación.
 Juan / the crisis has got better / worse the situation
"Juan / the crisis has improved/got worse the situation."

- (89) Jan / de wind bewoog de tafel.
 Jan / the wind moved the table
"Jan / the wind has moved the table."

My claim is that their basic entry is as in (90). They select an external argument [+c] (compatible with agentive and causative readings) and an internal theme [-c-m] argument. Optionally, they can also select a goal [-c] unitary argument.

- (90) $V_{acc} u[+c]_1, u[-c-m]_2 (u[-c])_2$

If the argument [-c] realizes, it has to be interpreted as [-c+m] (since there is a [c-m] argument already) due to the Full Interpretation of Thematic Roles principle in (91) (Marelj 2004). Hence, the animacy restriction.

- (91) Full Interpretation of Thematic Roles (FITR): *(Marelj 2004:67)*

For the purposes of interpretation, all custers must be fully specified.

If decausativization does not apply and the argument [-c] does not realize, there are two λ operators ([+c] and [-c-m]). If the argument [-c] realizes (interpreted as [-c+m] at C-I due to FITR), then there are three λ operators.

If decausativization applies and the argument [-c] does not realize, there is only one λ operator. If the argument [-c] does realize, there are two λ operators.

- (92) a. Jan_[+c] heeft (hem_[-c])_i; de tafel_[-c-m] bewogen (voor zijn opa)_i.
 Jan_[+c] has (him_[-c]) the table_[-c-m] moved (for his grandfather)
"The wind has moved the table (to this grandfather)."
- b. De tafel_[-c-m] heeft (*zich_[-c]) bewogen.
 The table_[-c-m] has (*zich_[-c]) moved
"The table has moved."

Note that in Dutch, when decausativization applies, *zich* is not required (as with TUVs) since PRO' is not spelled out. If *zich* is present in sentences such as (92), it is an anaphor that fulfil the DCC (27) and is interpreted as a semantic argument of the event denoted by the verb at the C-I system. In this case, *zich* is not the realization of PRO' but it is an anaphor that realizes the [-m] theta-cluster (interpreted as [-c+m] at C-I due to FITR).

When [-c] is realized, the argument that bears it can be bound by the theme [-c-m] argument if decausativization applies (or even by the [+c] argument without decausativization). This is possible because the argument [-c] bears inherent dative Case. Due to its inherent Case, the [c] argument does not need to form any chain with any other functional head (its has both its Tns- and theta-features valued and interpretable) and thus, the A-binding takes place at C-I. In other words, the binding relation between the [-c] argument and the [-c-m] or [+c] argument is not achieved by forming chains in the C_{HL} but by means of A-binding at the C-I system. Hence, no protector element MISMO is needed, because there is no chain formation in C_{HL} . I will call this kind of reflexive binding *pseudo-reflexive binding* (93) because, unlike standard reflexive-binding, it does not need a protector element due to the inherent Case of one bound element, which forces the binding to take place at C-I instead of at C_{HL} .

(93) Pseudo-reflexive binding:

- a. Binding between two arguments when the bound one bears inherent Case.
- b. Due to the inherent Case, the bound element does not need to form any chain in the syntax to be licensed, which forces the binding to take place by A-binding at C-I.
- c. Since the binding takes place at C-I and no chain is formed in C_{HL}, no protector element such as SELF or MISMO is necessary.

In Spanish, however, *se* is always required as the realization of PRO' (as with TUVs), and so the animacy restriction is obscured, as can be seen in (94).

- (94) a. [+c]El viento_i ([-c]le*_i/*_j/k)_{dative} ha movido [-c-m] la mesa_j.
 [+c]The wind_i ([-c]him*_i/*_j/k)_{dative} has moved [-c-m] the table_j
"The wind has moved the table (to him)."
- b. [-c-m]La mesa_i *(no θ / *[-c]se_{no index}/*_i)_{dative} ha movido.
 [-c-m]The table_i *(no θ / *[-c]se_{no index}/*_i)_{dative} has moved
"The table has moved."

6.5.1.2. Envejecer(se)-type verbs

This kind of verbs, such as (83) above, allow causative alternation as in (95a,b). They impose animacy restrictions, i.e. the pronominal particle is possible only with animate [-c-m] subjects as seen in (95c,d). Therefore, they are alternating pronominal verbs in Spanish as seen in (83).

- (95) a. [+c]El tiempo_i ([-c]le*_i/*_j/k=animate)_{dative} ha envejecido [-c-m] la mesa_j.
 [+c]The time_i ([-c]him*_i/*_j/k=animate)_{dative} has made older [-c-m] the table_j
"The time has made his table older."
- b. [+c]Las drogas_i ([-c]le*_i/j/k=animate)_{dative} han envejecido [-c-m] a Juan_j.
 [+c]The drugs_i ([-c]him*_i/j/k=animate)_{dative} have made older [-c-m] to Juan_j
"Drugs have made Juan older."
- c. [-c-m]La mesa_i (*[-c]se_i/j)_{dative} ha envejecido.
 [-c-m]The table_i (*[-c]se_i/j)_{dative} has got older
"The table has got older."
- d. [-c-m]Juan_i ([-c]se_i/*_j)_{dative} ha envejecido.
 [-c-m]Juan_i ([-c]se_i/*_j)_{dative} has got older
"Juan has got older."

Note that the dative *le* in (95a,b) is only possible with animate arguments since it is the [-c] argument, as it happens with *se* (also dative) in (95c,d) that realizes the [-c] theta-cluster.

My claim is that their basic entry is (96). These verbs select a [+c] external argument, a [+c-m] argument, a theme [-c-m] and an optional [-c] unitary cluster.

(96) $V_{acc} u[+c]_1, i[+c-m], u[-c-m]_2 (u[-c])_2$

Since there is a [+c] cluster, it can undergo *decausativization* and hence, the causative alternation shown in (83) and (95). The [+c-m] is conflated and hence, the lack of pronominal particle (i.e. the materialization of PRO') with decausativized entries as in (95c,d). As said before, the [-c] argument is realized by anaphors like *le* or *se*, which undergo *pseudoreflexive binding* since they are marked with dative case and the A-binding takes place directly at the C-I system due to the impossibility of establishing chains with dative arguments.

6.5.2. Verbs without animacy restrictions: theme-subject matter verbs

These verbs do not allow causative alternation nor do they show animacy restrictions as seen with the verbs *caer(se)* (80), *parar(se)* (82), *morir(se)* (81); and verbs such as *engordar(se)* and *adelgazar(se)* in South American Spanish.

We can see in (97) that these verbs do not allow a standard causative alternation. Sentences (97a) are only acceptable for some speakers¹¹, whereas (97b) are completely unacceptable. In (97c,d) we see that these verbs enter in a pseudocausative alternation, which indicates us that these sentences are the result of the application of the operation *causativization* that adds an agent [+c+m] to the basic verbal entry.

(97) a. %El viento / %Pedro cayó el vaso.
 %The wind / %Pedro fell the glass
 "The wind / Pedro made the glass fall."

¹¹ In some regions like Extremadura, the transitive use of this verb is allowed. However, the subject does not have any volitionality (i.e. it is /-m), if so, the verb *tirar* (throw) must be used, like in Standard Spanish.

- b. *El viento / *Juan murió a Ana.¹²
 *The wind / *Juan died to Ana
"The wind / Juan made Ana die"
- c. *La tormenta /*los frenos/el maquinista paró el tren. (causativized,
 cf. Reinhart 2002)
 *The storm /*the brakes/the engine driver stopped the train
"The storm / the brakes / the engine driver stopped the train."
- d. *El comer/*la comida/el granjero engorda los cerdos. (causativized,
 cf. Reinhart 2002)
 *The eating/ *the food /the farmer fattens the pigs
"Eating / the food / the farmer fattens the pigs."

My claim is that these verbs are theme-subject matter verbs for which their causative use has not been grammaticalized and only the decausativized entry is available. The operation is defined in (98).

(98) Frozen decausativized theme-subject matter verbs:

- a. Basic entry (not grammatic.): $V_{acc} u[+c]_1 i[+c-m] u[-c-m]_2 u[-m]_2$
- b. Causative use (not grammatic.): $V_{acc} u[+c]_1 i[+c-m] u[-c-m]_2 u[-m]_2$ (due to FITR)
- c. Decausativization: $R_d(V) i[+c-m] u[-c-m]_2 u[-m]_2$

The operation takes as input (which is not grammaticalized either) a verb with a cause [+c], an argument [+c-m], a theme [-c-m] and a unitary subject matter theta-cluster [-m] (98a). The [+c-m] argument is conflated and hence, a compound (NP+V) is formed with one interpretable [+c-m] theta-cluster. Due to FITR, the [-m] and the [+c] cannot co-realize and the argument [+c] is realized.

However, this entry (98b) is not grammaticalized and just the decausativized entry (98c) can be realized, i.e. the [+c] argument is deleted altogether due to the application of *decausativization*. Then we have a compound (NP+V) with an interpretable [+c-m] theta cluster and two uninterpretable theta-clusters: a theme [-c-m] and a subject matter [-m]. The [-c-m] and [-m] arguments cannot co-realize due to FITR and restrictions on θ -cluster co-realization (Marelj 2004), as specified in (99) below.

¹² In Catalan some residual transitive uses of the verb *morir* are allowed like *"M'has mort"* (You have killed me) (Gallardo 2008). This use is very restricted and the subject has animacy restrictions, which points towards the possibility of this used being derived by a causativization (Reinhart 2002) operation in the lexicon, which always adds agents [+c+m] arguments to the verbal entry.

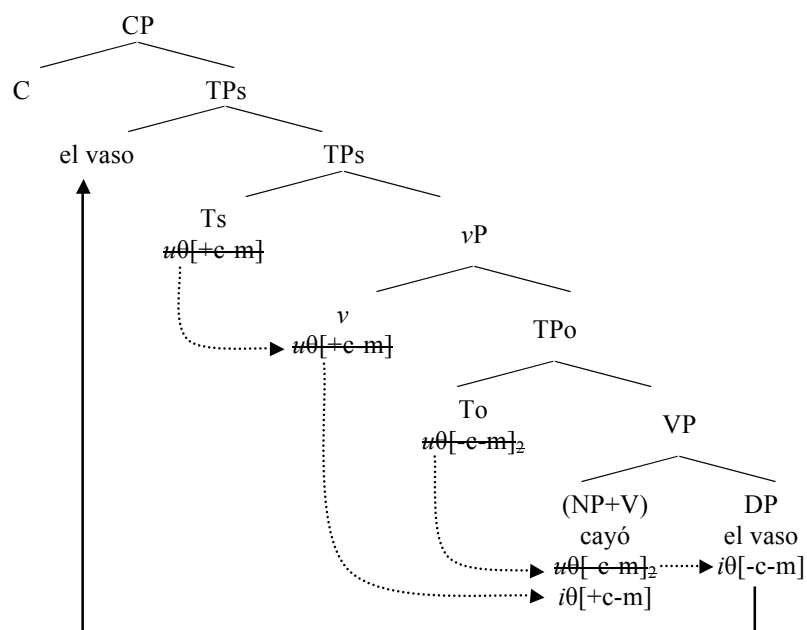
- (99) a. $i[+c-m] u[-c-m]_2 u[-m]_2 \Rightarrow i[+c-m] u[-c-m]_2 u[-c-m]_2 \Rightarrow$ FITR violation
 b. $i[+c-m] u[-c-m]_2 u[-m]_2 \Rightarrow i[+c-m] u[-c-m]_2 u[+c-m]_2 \Rightarrow$ FITR violation

If [-c-m] realizes, no pronominal particle is needed (the θ -feature on both Ts and To can be deleted).

(100) El vaso cayó.

The glass fell

"The glass fell off"



However, if [-m] realizes, the pronominal particle needs to be inserted in order to delete the θ -feature on To due to the fact that the DP fulfils the *Activity Condition* (103) as described in (104). The relation between To and DP_{object} is mediated by V; since V lacks the /c feature on the cluster [-m], the θ -chain does not contain any /c feature, and the /c feature on To cannot be deleted). The head To cannot "see" DP_{object} because it has inherent Case and hence it is *inactive* for further computations in the syntactic derivation due to the *Activity Condition* (102) redefined in our terms in (103).

(101) **Inactivity of an XP:** An XP that eliminates its uninterpretable features (case, *wh*) is rendered *inactive*. (Nevins 2004:9)

(102) **The Activity Condition:** Inactive elements are not accessible for further operations. (Nevins 2004:9)

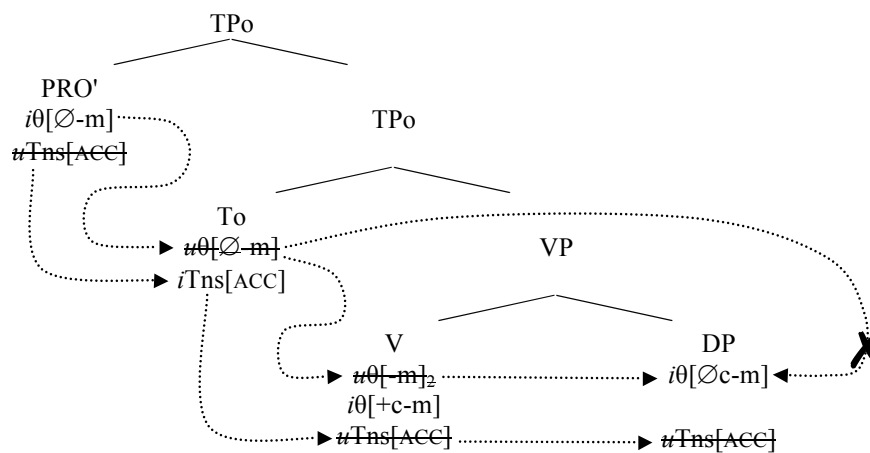
(103) The Activity Condition (reformulated in terms of agree-chains):

A nominal argument (DP) is no longer active for establishing agree chains in the syntactic derivation iff:

- a. has its Tns-features (Case) valued and in a Tns-chain with a functional head with an interpretable instance of the Tns feature or has inherent Case (so that the uninterpretable instance on the nominal can be deleted),
- b. and has its theta-features valued by virtue of a theta-chain with a functional head (V, P)

The agree relation between V and To in (104) is not enough to delete the interpretable θ -feature /c on To since the verb does not have this feature (its θ -cluster is unitary), and thus it cannot establish a chain with To to share this feature and make it deletable on To. Moreover, To cannot establish a relation with DP complement of the verb in (104) because it has all its Tns- and θ -features valued or form part of a chain which contains an interpretable instance for all its features (Inactivity Condition, based on Chomsky 2001: if a DP has all its Tns- and θ -feature valued, then it does not need to form any other chain with any other functional category and it is "inert" for further agree-relations in the syntactic derivation, this is precisely what happens with DPs marked with inherent Case). PRO' is inserted as a last resort mechanism in [Spec,TPo] so that it deletes the θ -feature /c on To. As result of this, it values its Tns, θ - and person φ -features, which takes the value of the DP complement of the verb and so it gets spelled out.

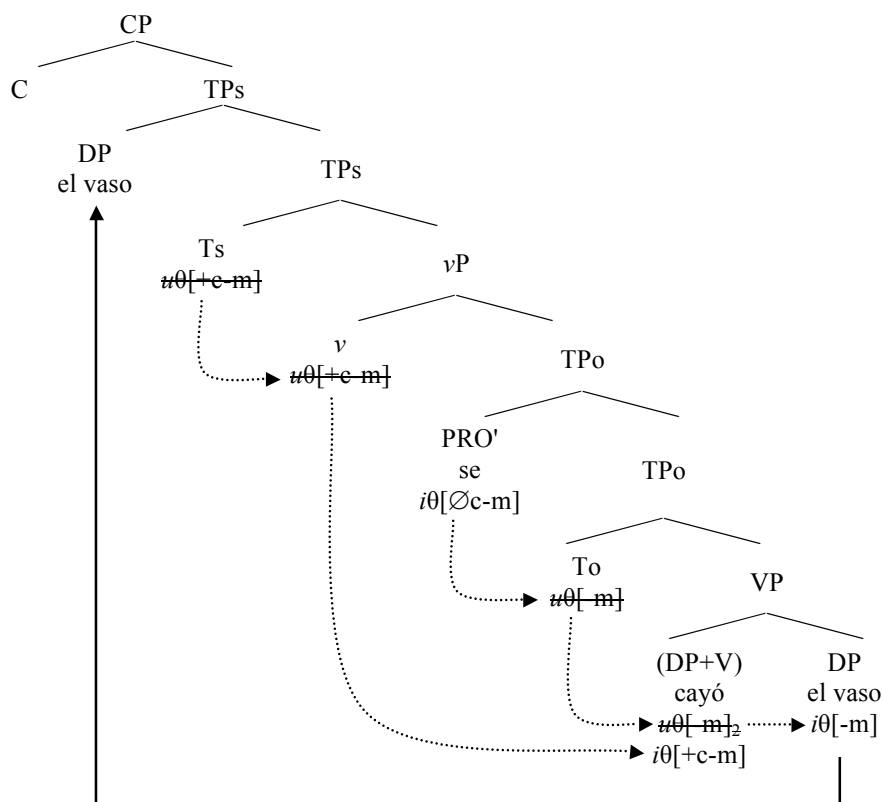
(104) The Inactivity Condition and [-m] arguments:



(105) El vaso **se** cayó.

The glass **se** fell

"The glass fell off."



The relation between PRO' and To is direct and not mediated with V (unlike with TUV, where V mediates the relation between PRO' and To). Finally, PRO' is materialized in Dutch iff its person feature is valued and it establishes a probe-goal relation with V. The SE-anaphors spell-out rules are redefined as in (106) below.

(106) SE-anaphors spell-out rules (revisited):

- Spanish spells out the ϕ -feature (person) of PRO' iff it is valued.
- Dutch spells out the ϕ -feature (person) of PRO' iff it is valued and PRO' establishes a Tns-chain with To and V (or via V).
- English never spells out the ϕ -feature (person) of PRO'.

6.6. Aspectual *se*

Aspectual datives like (107) and (108) are derived from the *causativization* operation (109) on agentive verbs, more concretely on creation and consumption verbs (though it is not operative on agentive verbs that allow the application of *reflexivization*).

(107) a. Juan **se** comió la pizza.

Juan **se** ate the pizza

"Juan ate up the pizza"

b. Juan comió la pizza.

Juan ate the pizza

"Juan ate the pizza"

(108) a. Juan **se** leyó el libro.

Juan **se** read the book

"Juan read the book (and he finished it)"

b. Juan leyó el libro.

Juan read the book

"Juan read the book"

The operation *causativization*, formalized in (109) takes as input a transitive verbal entry (109a), and it adds an agent [+c+m] argument (109b) so that the first [+c+m] theta-cluster needs to be decausativized into a [-c+m] argument. The causativization is an expansion operation that augments the verb's arity from two arguments to three ones, both in syntax (109c) and semantics (109d).

(109) Causativization

(Reinhart 2002, Horvath & Siloni 2008)

a. Basic entry (transitive): $V (u[+c+m]_1, u[-c-m]_2)$

b. Causativization: $E_c(V)_{acc} (u[+c+m]_1, u[-c+m], u[-c-m]_2)$

c. Syntactic realization: $DP_{[+c+m]1} vP DP_{[-c-m]2} DP_{[-c+m]}$

d. Interpretation: $\exists e\lambda x\lambda y\lambda z [e=V \& [+c+m], e=x \& [-c+m], e=y \& [-c+m], e=z]$

The aspectual dative [-c+m] is an argument (unlike the pronominal particle in intransitive verbs), i.e. it fulfills de DCC (27) and hence it is visible at the C-I system, and interpreted as a semantic argument as seen in (109d).

The aspectual dative realizes the [-c+m] cluster. Note that it is [+R] in the following examples:

(110) a. Juan_[+c+m] **se**_[-c+m] leyó el libro_[-c-m].

Juan_[+c+m] **se**_[-c+m] read the book_[-c-m]

"Juan read the book (and this brought about a change of state in Juan)."

- b. Juan_[+c+m] **le**_[-c+m] leyó el libro_[-c-m].
 Juan_[+c+m] **him**_[-c+m] read the book_[-c-m].
"Juan read the book to him."
- c. Juan_[+c+m] **me**_[-c+m] leyó el libro_[-c-m].
 Juan_[+c+m] **me**_[-c+m] read the book_[-c-m].
"Juan read the book to me."
- d. Juan_[+c+m] **nos**_[-c+m] leyó el libro_[-c-m].
 Juan_[+c+m] **us**_[-c+m] read the book_[-c-m].
"Juan read the book to us."

Note that even a full DP can realize the aspectual *se*, as in (111), given that it is an argument with a theta-cluster and fulfils the DCC (27).

- (111) Juan_[+c+m] **le**ⁱ_[-c+m] leyó el libro_[-c-m] a Anaⁱ.
 Juan_[+c+m] **her**ⁱ_[-c+m] read the book_[-c-m] to Anaⁱ.
"Juan read the book to Ana."

Note that in (110a), *se* seems to be locally A-bound by the antecedent *Juan*, rendering a reflexive reading. This is not the case, however. Note in (109d) that there is no reflexive interpretation. There are two lambda operators that bind two different variables realizing two different θ -clusters. In (110a) there is indeed binding of *se* by *Juan*, but this is done non-locally since *se* realizes [-c] and inherent dative Case. Therefore, the binding relation is done at C-I (otherwise, the Theta Criterion would be violated as in the case of intransitive pronominal verbs), in other words, it is a case of *pseudo-reflexive binding* defined in (93) above.

The aspectual differences that are observed in (107) and (108) are epiphenomenal: they are due to the [-c+m] experiencer theta-cluster that implies that the agent [+c+m] argument, which pseudo-reflexive binds (93) the [-c+m] argument, experiences a (minimal) change of state by virtue of the event denoted by the verb. In other words, the aspectual differences are consequence of the thematic contribution of the aspectual *se*. In (107a) the implication is that Juan ate the pizza up because Juan is somehow mentally involved in the event of eating (*se* bears the experiencer [-c+m] theta-cluster and it is bound by *Juan*), so he has to eat enough quantity of pizza in order to experience the consequences of this event of eating (due to the [-c+m] theta-cluster).

Therefore, *Sócrates* in (112) below has to drink enough amount of venom so as to experience a change of state (die) although he does not need to drink all the venom¹³.

(112) *Sócrates se tomó veneno.* (Romero 2009)

Sócrates se drank venom

"Sócrates drank venom."

In (108) this difference is even clearer: in (108a) it is implied that Juan read the book and understood what he read, because this caused a change in his mental state regarding the event of reading. In (108a) this is not implied, so Juan could have read the book aloud but without understanding a word of it. This last implication is not obtained in (108a).

6.7. Complex [+] entries

These are verbs like those in (113a) and (114a), which can undergo both decausativization (115b) and (116b), and reflexivization (115) and (116c), i.e. they are pronominal verbs that are ambiguous: they can be either inchoative (113b) and (114b), or reflexive (113c) and (114c).

(113) a. *Juan se cortó.*

Juan se cut

"Juan cut himself." / "Juan was cut."

b. $\exists e\lambda x [e=V \ \& \ [-c-m],e=x]$

c. $\exists e\lambda x [e=V \ \& \ [+c+m],e=x \ \& \ [-c-m],e=x]$

(114) a. *Juan se mató.*

Juan se killed

"Juan killed himself." / "Juan was killed."

b. $\exists e\lambda x [e=V \ \& \ [-c-m],e=x]$

c. $\exists e\lambda x [e=V \ \& \ [+c+m],e=x \ \& \ [-c-m],e=x]$

The lexical entry of these verbs has both a cause [+c] and an agent argument [+c+m] as in (115a) and (116a). They cannot co-realize due to their being [+] clusters, and depending on which θ -cluster is realized, either decausativization (115b) and (116b), or reflexivization (115c) and (116c) applies.

¹³ See Romero (2009) for a similar proposal based on other grounds.

- (115) a. $V(\text{cortar})_{\text{acc}} [+c]_1, [+c+m]_1, [-c-m]_2$
 b. $R_d(\text{cortar}) [-c-m]_2$
 c. $R_r(\text{cortar}) ([+c+m], [-c-m])_1$
- (116) a. $V(\text{matar})_{\text{acc}} [+c]_1, [+c+m]_1, [-c-m]_2$
 b. $R_d(\text{matar}) [-c-m]_2$
 c. $R_r(\text{matar}) ([+c+m], [-c-m])_1$

6.8. Conclusions

We have seen in this chapter why the pronominal verbs require the pronominal particle, what this particle is, and why the non-pronominal verbs do not allow its presence. Besides, we have accounted for the (pseudo-)optionality of the pronominal particle in alternating pronominal verbs, and its aspectual role with consumption and agentive verbs. Finally we have accounted for the ambiguity between the inchoative and the reflexive reading with verbs such as *cortar* (cut) and *matar* (kill).

The pronominal verbs are inherent reflexive verbs (IRVs), subject experiencer verbs (SEVs) and theme unaccusative verbs (TUVs). All these verbs undergo a reduction operation (either *reflexivization* or *decausativization*) by which their arity is reduced in one argument. The syntax always requires the merging of two nominal elements, though, in order to delete the uninterpretable theta-features on the tempo-aspectual heads of the sentence T_s and T_o . Therefore, a null minimally specified SE-anaphor PRO' is merged either in object or in subject position so that the uninterpretable theta-features that remain unvalued enter into an agree relation with the interpretable instances of PRO' and hence, can be deleted so that the derivation converges at the interface with the C-I system. Depending on the language, the person-feature of PRO' (which is valued with the person feature of the argument introduced by the verb) is spelled out or not.

The non-pronominal verbs do not require (and do not allow) the pronominal particle because they have undergone a *conflation* operation, which conflates or incorporates a null NP, by which one uninterpretable θ -cluster becomes interpretable. Therefore, the interpretable theta-cluster deletes the uninterpretable theta-features of either T_s or T_o , and the argument introduced by the verb deletes the uninterpretable theta-cluster of the other tempo-aspectual head.

The alternating pronominal verbs, which can impose animacy restrictions or not, are basically accounted for by the presence of a unitary theta-cluster that can be either an optional goal [-c] or a subject matter [-m].

The aspectual *se* that optionally appears with some agentive verbs, like consumption verbs, are [-c+m] arguments result of the application of the operation *causativization*, which is an expansion operation that augments the verb's arity in one argument.

Finally, the verbs like *cortar* (cut) and *matar* (kill) are ambiguous between an inchoative and a reflexive reading because they have two [+c] arguments that cannot co-realize: they a cause [+c] theta cluster and an agent [+c+m] theta cluster. If [+c] is realized, decausativization can apply and hence, the inchoative reading. If the [+c+m] realizes, reflexivization can apply and hence, the reflexive reading.

Table (118) summarizes the lexical operations studied so far (reduction, reflexivization and causativization) and the new lexical operation that has been defined in this chapter (conflation).

(118) Operation	Input	Output	Based on
Reflexivization	[+c+m]	deletion of [+c+m]	Reinhart & Siloni (2005)
Reduction	[+c]	deletion of [+c]	Reinhart (2002)
Conflation	[/c-m]	conflation or [/c-m]	Hale & Keyser (2000)
Causativization	transitive entry	adds one [+c+m] argument	Reinhart (2002) Horvath & Siloni (2008)

The theta-composition of the different verbs so far studied, is summarized in (117) and (118) below.

(117) Intransitive verbs (Spanish & Dutch):

1. NON-ALTERNATING PRONOMINAL VERBS
 - 1.a. Inherent Reflexive Verbs: [+c+m]₁[-c-m]₂
 - 1.b. (Decausat.) Subject Experiencer Verbs: [+c]₁[-c+m]₁([-m]₂)
 - 1.c. (Decausat.) Theme Unaccusative Verbs (Spanish): [+c]₁[-c-m]₂
 - 1.d. Complex [+] Verbs: [+c]₁[+c+m]₁[-c-m]₂

2. ALTERNATING PRONOMINAL VERBS

2.a.[+c][-c] verbs:

2.a.a.*Mejorar(se)* / (*zich*) *bewegen*-type verbs [+c]₁[-c-m]₂([-c]₂)

2.a.b.*Envejecer(se)* verbs..... [+c]₁[+c-m][[-c-m]₂([-c]₂)

2.b.Theme-subject matter verbs: [+c-m][[-c-m]₂[-m]₂

3. NON-PRONOMINAL VERBS

3.a.(Decausat.) Theme Unaccusative Verbs (Dutch): [+c]₁[-c-m]₂

3.b.Internally caused unaccusative verbs *floreecer*-type: [+c-m][[-c-m]₂

3.c.Internally caused unaccusative verbs *augmentar*-type:..... [+c+m]₁[+c-m][[-c-m]₂

3.d.Existential and presentational unaccusative verbs: [+c-m][[-c-m]₂[-c]₂

3.e.Double caused verbs: *cambiar*: [+c][+c-m][[-c-m]₂

(118) Transitive pronominal verbs (Spanish):

4. ASPECTUAL DATIVES WITH AGENTIVE VERBS:..... [+c+m]₁[-c+m][[-c-m]₂

Finally, we can see in the table (119) the full catalogue of anaphors available in UG.

(119) PRO – Pronominals Continuum:

	SE-anaphors			SELF-anaphors	pronominals	
	PRO'	PRO	<i>se/sí</i>	<i>zich</i>	himself zichzelf	he/him/his aspectual datives
R	[-R]	[+R]	[-R]	[-R]	[-R]	[+R]
Φ-features	<i>i</i> φ person [-]	<i>i</i> φ person [-] <i>i</i> φ number [-]	<i>i</i> φ person [3 ^a]	<i>i</i> φ person [3 ^a]	<i>i</i> φ person [3 ^a] <i>i</i> φ gender [♂] <i>i</i> φ number [sing]	<i>i</i> φ person [3 ^a] <i>i</i> φ gender [♂] <i>i</i> φ number [sing]
Structural Case	<i>u</i> T [-]	<i>u</i> T [-]	<i>u</i> T [-]	<i>u</i> T [-]	<i>u</i> T [-]	<i>u</i> T [-]
Tns-chain with	-	Ts or To	Ts or To	To	Ts or To	Ts or To
Phonological content	no	no	yes	yes	yes	yes
Binding at C _{LH}	yes	yes	yes	yes	no	no
Reflexivizing function	no	no	no	no	yes	no

The most simple SE-anaphor is PRO', without phonological content and with just a person feature. Since it lacks grammatical number, it is [-R]. PRO, on the other hand, is specified with grammatical number since in certain contexts, can be [+R]. The SE-

anaphors *se*, *sí* (Spanish) and *zich* (Dutch) are the materialized instances of PRO' in certain contexts, depending on the language under study. On the other hand, there are SELF-anaphors that can force a reflexive reading fulfilling R&R's Condition B. Finally, there are pronominals, that are [+R] and cannot license a reflexive reading.

Chapter 7

Recapitulation and implications: SE-anaphors in arbitrary constructions

7.1. Recapitulation

So far we have seen that a series of syntactic constructions are related by the presence of a SE-anaphor, be this null (PRO or PRO') or overt (*zich, se, si*, etc.). Table (1) presents a summary of the data in the different languages so far studied and conclusions so far reached:

(1) Recapitulation:

	English	Dutch	Romance	Slavic
OC	PRO \Rightarrow \emptyset	PRO \Rightarrow \emptyset	PRO \Rightarrow \emptyset	PRO \Rightarrow \emptyset
NOC	PRO \Rightarrow \emptyset	PRO \Rightarrow \emptyset	PRO \Rightarrow \emptyset	PRO \Rightarrow \emptyset
Reflexive	PRO' \Rightarrow \emptyset	PRO' \Rightarrow <i>zich</i>	PRO' \Rightarrow <i>se/si</i>	PRO' \Rightarrow <i>się</i>
Ergative	PRO' \Rightarrow \emptyset	PRO' \Rightarrow \emptyset	PRO' \Rightarrow <i>se/si</i>	PRO' \Rightarrow <i>się</i>
Inherent	PRO' \Rightarrow \emptyset	PRO' \Rightarrow <i>zich</i>	PRO' \Rightarrow <i>se/si</i>	PRO' \Rightarrow <i>się</i>
Aspectual	verbal particles	verbal particles	PRO' \Rightarrow <i>se/si</i>	?

In chapter 3 we have seen that both obligatory control and non-obligatory control have different syntactic characteristics and semantic interpretation of the null SE-anaphor PRO, which has a person feature as well as a grammatical number feature that allows it to be referentially independent ([+R]), as in cases of non-obligatory control. In cases of obligatory control, PRO is locally bound by its antecedent by means of syntactic chains.

In chapter 4 we have reviewed the main literature on Romance *se/si* clitics and analogous elements in Slavic and Germanic languages. We have focused on both the status and the function of the clitics in the different constructions where they appear.

Reflexive verbs have been studied in chapter 5. We have distinguished inherent reflexive verbs from non-inherent reflexive verbs. The former need an overt SE-anaphor due to conditions imposed by the syntax (an uninterpretable instance of theta has to be deleted by the SE-anaphor), whereas the latter require the presence of a SELF-anaphor, which are composed of a pronoun or a SE-anaphor plus a protector element. The SE-

anaphors that appear with inherent reflexive verbs are materialized in Romance languages (*se/si* clitics) and with some Germanic languages such as Dutch (*zich*) but not with others like English, which uses zero-morphology. This depends on rules of SE-anaphors materialization on the syntax-phonology interface.

Finally, we have seen in chapter 6 that ergative and inherent *se/si* (what we call *pronominal verbs*) is an instance of PRO' which has been materialized and is inserted due to conditions imposed by the syntax: again and in a parallel way to reflexives, the uninterpretable instance of theta on either the aspectual head within the vP or the tense head of the sentence has to be deleted by agreeing with this null SE-anaphor. The variation of the occurrence of the materialized clitic in Romance, or particle (*zich*) in Dutch or zero-morphology in English is due to different conditions that affect the rules of materialization of SE-anaphors that apply at the syntax-phonology interface.

This chapter is mainly devoted to reviewing previous analyses of the clitic *se/si* in Romance in impersonal (2) and passive constructions (3). At the end of the chapter I will sketch a new analysis of these constructions (Romance impersonal and passive *se/si*) as well as of arbitrary control across languages (4).

(2) a. **Se** baila mucho en las fiestas. (Mendikoetxea 2008:303)

Se dance_{singular} a lot in the parties

"One (SE) dances a lot at parties."

b. Gli **si** telefona spesso. (Burzio 1986:43)

Him_{dative} **si** phone_{singular} often

"We phone him often."

(3) a. **Se** comen las manzanas. (Mendikoetxea 2008:291)

Se eat_{plural} the apples

"The apples are eaten."

b. **Si** mangiano le mele. (Mendikoetxea 2008:291)

Si eat_{plural} the apples

"The apples are eaten."

(4) a. [**PRO**_{arb} reading the Silmarillion] is difficult.

b. [Het **PRO**_{arb} lezen van het Silmarillion] is moeilijk.

[The **PRO** read of the Silmarillion] is difficult.

"Reading the Silmarillion is difficult."

- c. [**PRO**_{arb} Leer el Silmarillion] es difícil.
 [**PRO** read the Silmarillion] is difficult.
"Reading the Silmarillion is difficult."

The hypothesis is that in all these constructions, there is a SE-anaphor in subject position: either PRO (in arbitrary control) or PRO' (in impersonal and passive *se/si*). I will also argue that *zich* in subject position in matrix clauses is not possible in Dutch since this language does not materialize PRO' if it agrees with Ts (it has nominative). The table (1) above would be completed with the analysis of the aforementioned constructions in table (5) below.

(5) Arbitrary constructions:

	English	Dutch	Romance	Slavic
AC	PRO \Rightarrow \emptyset	PRO \Rightarrow \emptyset	PRO \Rightarrow \emptyset	PRO \Rightarrow \emptyset
Impersonal <i>se/si</i>	does not exist	does not exist	PRO' \Rightarrow <i>se/si</i>	PRO' \Rightarrow się
Passive <i>se/si</i>	does not exist	does not exist	PRO' \Rightarrow <i>se/si</i>	PRO' \Rightarrow się

To sum up, in this chapter I will analyse the role of the SE-anaphors in arbitrary constructions in order to complete the possible scenarios where this kind of anaphors can appear (with the exception of middles, which will be left open for future research).

7.2. Arbitrary constructions

7.2.1. Arbitrary control¹

Arbitrary control takes place when PRO cannot be controlled either by a local antecedent (OC) or a non-local antecedent (NOC), as in (6b).

- (6) a. Hi **John**_i. What are you_i doing? **PRO**_{i/*j} reading that book?
 b. **PRO**_{arb/*j} reading that book is difficult
- (7) **PRO**_{i/*j} reading that book is difficult for **me**_i

When PRO cannot be controlled because there is no antecedent whatsoever, not even in the discourse storage, it takes *arb* value as a last resort. This is the most costly option: *arb* is blocked in (6a) because there is a possible controller and this may be done by covaluation at discourse level in the same way as in (7). In (6b) there is, however, no

¹ See Manzini & Roussou (2000:427).

other option than *arb* (assuming that the sentence is completely isolated or the context provides no suitable antecedent). The true value *arb* is only compatible with a totally free pronoun (i.e. not constrained by Φ -features) since *arb* may refer to speaker, addressee/s, other/s or any possible combination. If the pronoun had any Φ -feature specified², its interpretive freedom would be constrained. Hence *arb* is only compatible with PRO (and with PRO' as will be seen later on).

7.2.2. Impersonal and passive SE

In chapter 4 I reviewed the most important analyses of impersonal (8) and passive (9) *se/si* in Romance, which can roughly be divided in two groups³. The first is composed of those authors who think that the *se/si* clitic is an argument or forms a chain with an argument, and as such, is subject to the Theta Criterion (Belletti 1982; Burzio 1986; Manzini 1983; Cinque 1988 for the cases of [+arg] *si*, Dobrovie-Sorin 1998 for impersonal *se/si*, Raposo & Uriagereka 1996). On the other hand there are other authors for whom the clitic *se* is not an argument but marks something on the INFL head (Otero 1986; Cinque 1988 for the cases of [-arg] *si*, Mendikoetxea 1992, Dobrovie-Sorin 1998 for passive *se/si*).

- (8) a. **Se vende** libros. (Mendikoetxea 1999:1676)
Se sell_{singular} libros
"One sells books."
- b. **Si leggerà** volentieri (alcuni articoli). (Burzio 1986:43)
Si will read_{sing} willingly (a few articles)
"We will be eager to read a few articles."
- (9) a. **Se pasaron los trabajos a** ordenador. (Mendikoetxea 1999:1635)
Se wrote the papers to computer
"The papers were typed in the computer."

² Dutch *men* has an *arb* value but I do not consider it a real *arb*. In sentences like (I), *men* usually refers to a group of people where neither the speaker nor the addressee are usually included unless otherwise indicated. This must be due to its person feature specified as 3rd (which can be seen in the agreement morpheme of the verb):

(i) Men zegt dat jij een leugenaar bent
 men says_{3rd,sing} that you a liar are

³ See chapter 4 section 4.7 for a complete overview of the studies realized on the different uses of *si/se* across languages.

b. **Si** leggeranno volentieri (alcuni articoli). (Burzio 1986:43)

Si will read_{plural} willingly a few articles

"A few articles will be read eagerly."

The hypothesis that I will further develop in this chapter is that in both impersonal and passive *si* constructions, the clitic *si/se* is the spell-out of a defective anaphor in subject position (in fact, I will argue it is a SE-anaphor in the sense of Reinhart & Reuland 1993 and Reuland 2001⁴). Being this anaphor [-R], i.e. it cannot be referential by itself; it is semantically interpreted by means of a choice function that ranges from existential to universal readings. Arbitrary control will be accounted for on similar grounds.

7.2.3. The semantics of arbitrary: Review

In this section I will briefly review three proposals on the semantics of arbitrary constructions, and particularly on the semantics of arbitrary *si/se*. First, I will present Cinque's (1988) Theory of Arb. I will move on to Chierchia's (1995) analysis based on existential disclosure, and I will finish the review with Mendikoetxea's (2002, 2008) proposals that conclude that aspect and arbitrary vs. generic interpretation are independent from one another.

Cinque (1988:542) distinguished three semantic interpretations of *si* in Italian finite clauses:

1. In generic sentences, a generic (arbitrary) interpretation is available, which is roughly paraphrasable as "people" or "one".
2. In sentences with specific time reference:
 - a. *Si* retains its generic/arbitrary reading only when it appears with transitive and unergative verbs.
 - b. With all the other verb classes (ergative, psych, copula, passive and raising verbs), *si* acquires a new interpretation roughly paraphrasable as "unspecified set of people including the speaker" or "we".

Cinque developed in this Theory of Arb an account of the data he found, based on his distinction between [\pm arg] *si*. Besides *si*, the Theory of Arb gives a unified explanation of other elements that receive an arbitrary interpretation, such as PRO_{arb}, French *on*, 2nd

⁴ Like Dobrovie-Sorin (1998, 2006) and Rivero (2001) do.

person singular pronoun, 3rd person plural pronoun, the restricted instances of *arb* in non [NP, IP] position such as object *pro_{arb}* (Rizzi 1986) and PRO_{arb}.

Cinque's Theory of Arb identifies an *arb* abstract element that has two usages, which differ systematically in a number of ways, and which he calls the **quasi-existential** and **quasi-universal** usages. Their different properties are summarized in (10). These are just two variants of one and the same *arb* and one important difference between both usages is that the quasi-universal usage is apparently not constrained by theta-requirements unlike the quasi-existential one.

(10) Theory of Arb: (Cinque 1988:546)

<i>Quasi-existential Interpretation</i>	<i>Quasi-universal Interpretation</i>
a. Compatible with specific time reference.	a'. Incompatible with specific time reference.
b. Incompatible with generic time reference.	b'. Compatible with generic time reference.
c. Incompatible with contexts suspending the specificity of the time reference.	c'. Compatible with contexts suspending the specificity of the time reference.
d. Compatible with the existence of a single individual satisfying the description.	d'. Incompatible with the existence of single individual satisfying the description.
e. Restricted to [NP, IP] theta-marked in D-Structure.	e'. Not restricted to [NP, IP] theta-marked in D-Structure.

The claim is that *arb* subjects acquire the two different interpretations of quasi-universal and quasi-existential quantification as a function of the different time reference (generic and specific, respectively) of the tense/aspect of the sentence they appear in. In other words, they can be seen as two contextual variants of a single *arb* entity.

Under this assumption, properties (10a-d) of *arb* interpretation follow as simple consequences of this primitive difference.

As for property (10e), Cinque simply assumes that the quasi-existential interpretation of *arb* needs to be matched with Infl at D-Structure like Otero's (1986) mechanism of absorption of certain features by Infl. Given such assumption, the availability of this kind of interpretation with only transitive and unergative subjects follows directly.

Chierchia (1995) analyses the quantificational and anaphoric properties of impersonal *si* in Italian, summarized in (11), in the framework of Discourse Representation Theory. He follows previous syntactic analyses (Burzio 1986, Cinque 1988⁵) of impersonal *si* as being a subject clitic that expresses a non-specific, generally plural, human subject. The framework he uses for his semantic theory of the interpretation of impersonal *si* is, as said before, Discourse Representation Theory (DRT) but with some modifications⁶.

(11) Quantificational and anaphoric properties of impersonal *si*: (Chierchia 1995:108-110)

- A. Quantificational variability in generic vs. episodic sentences, as can be seen in (12). This was already noted by Cinque (1988), as well as the fact that this behaviour of *si* seems parallel to that of bare plurals in English (13). However, Chierchia notes that this parallelism breaks down in contexts with kind-level predicates, where English bare plurals are allowed but not Italian *si* (14).
- B. *Si* cannot antecede overt pronominals and *pro* as in (15a) and (15b) respectively.
- C. *Si* can antecede reflexives both clitic (16b) and non-clitics (16a) reflexives.
- D. *Si* can antecede PRO as in (17).
- E. *Si* can antecede itself as in (18).
- F. Variability of the quantificational force with if/when-clauses as in (19). The quantificational force of *si* appears to be determined by the quantificational adverb⁷, which parallels what happens with indefinites in conditional sentences.

(12) a. In Italia **si** beve molto vino. (Chierchia 1995:108)
 In Italia **si** drinks a lot of wine
"In Italy, everybody/people drink a lot of wine"

⁵ See chapter 4 section 4.2.

⁶ Chierchia himself (1995:119) summarizes the modified version of DRT that he adopts in the following assumptions:

- a. The scope of a quantificational adverb is what it is adjoined to at LF. Its restriction is what is external to the scope (and in the same local environment as the adverb).
- b. Indefinites are existentially quantified (contra Heim 1982).
- c. Adverbs of quantification "disclose" the indefinites they are coindexed with, where indices on quantificational adverbs are assigned freely, subject to standard conditions on binding. Hence, the *novelty condition* to which indefinites are subjects can be reduced to Chomsky's (1981) Condition C.

⁷ Conditionals are thought to have a null generic quantificational adverb.

- b. In Italia ieri **si** è giocato male. (Chierchia 1995:108)
 In Italia yesterday **si** played poorly
"Yesterday people in Italy played poorly"
- (13) a. Bears like honey. (Chierchia 1995:108)
 b. Bears are hibernating in this area. (Chierchia 1995:108)
- (14) a. On normal hiking trails, one meets a lot of people. But in the desert people are pretty rare. (Chierchia 1995:108)
 b. *Nel deserto **si** è piuttosto rari. (Chierchia 1995:108)
 In the desert **si** is rather rare
- (15) a. ***Si**_j è detto che loro_j hanno sbagliato. (Chierchia 1995:109)
Si said that they were wrong
 b. ***Si**_j è detto che pro_j vinceranno. (Chierchia 1995:109)
Si said that pro will win
- (16) a. **Si**_j è troppo spesso ingiustificatamente orgogliosi de se' stessij. (Chierchia 1995:109)
"People are too often proud of themselves for no reason."
 b. Ci **si** è lavati. (Chierchia 1995:109)
 Self_{clitic} **si** is washed
"People washed themselves."
- (17) **Si**_j è cercato di [PRO_j vincere]. (Chierchia 1995:109)
"People tried to win."
- (18) a. Se **si** gioca male, **si** perde. (Chierchia 1995:109)
"If people play badly, they lose."
 b. Ieri, **si** è giocato male e **si** è perso. (Chierchia 1995:109)
"Yesterday, people played badly and they/people lost."
- (19) a. Se **si** è alti **si** è sempre belli. (Chierchia 1995:110)
"Everyone who is tall is beautiful."
 b. Se **si** è alti, **si** è talvolta anche belli. (Chierchia 1995:110)
"Some people who are tall are also handsome."

From properties (11a) and (11f) Chierchia concludes that *si* behaves as an indefinite. Nonetheless, properties (11b-e) show that *si* also behaves as a definite (pronoun). Furthermore, if *si* is an indefinite, it should be subject to the novelty condition (to which

the indefinites are thought to be subject) and hence, the occurrence of *si* in the main clause of (19a) and (19b) should not be able to be anaphoric to the occurrence of *si* in the if-clause. This leads Chierchia to say that while *si* truly acts as an indefinite in the restriction of quantified structures, it seems to act as a pronoun (definite) in the scope of such structures.

The hypothesis that Chierchia puts forward in order to account for the properties in (11) claims that *si* is interpreted as an operation that takes a property and does two things⁸: first, it closes existentially the argument corresponding to the subject, and second, it restricts the range of such an argument to groups of humans (perhaps drawn from a contextually specified set), which is implemented by choosing a variable x_{arb} whose range is restricted to humans, being this formalized by means of a distinguished *arb* index in the variable introduced by *si*. *Si* can be thus regarded as a function from properties into formulae and define a general functor SI in (20) of the appropriate type, which will constitute the interpretation of *si*.

$$(20) \text{ SI}(P) = \exists x_{arb} [P(x_{arb})] \quad (\text{Chierchia 1995:121})$$

The quantificational variability in generic versus episodic sentences (property (11a)) follows from the fact that *si* existentially closes the subject argument, and the interaction of the existential operator introduced by *si* with other operators in the sentence. A sentence like (21a) is expected to be ambiguous depending on whether it is understood generically or episodically, being this ambiguity formally represented by the absence or presence of a generic operator⁹ in LF.

$$(21) \text{ a. Si canta.} \quad (\text{Chierchia 1995:120})$$

Si sings

"People sing."

$$\text{b. } \exists x_{arb} [\text{sing}(x_{arb})] \quad (x_{arb} = \text{variable restricted to ranging over groups of humans})$$

$$\text{c. Gn s}[C(s)] [\exists x_{arb} \text{ sing}(x_{arb})] \quad (x_{arb} = \text{idem})$$

$$\text{d. } \exists x_{arb} \text{ Gn s}[C(s, x_{arb})] [\text{sing}(s, x_{arb})] \quad (x_{arb} = \text{idem})$$

⁸ Note that *si* is interpreted as an *operation on properties* rather than being assigned an ordinary NP meaning (i.e. a *generalized quantifier*). Chierchia argues that this fits well with the fact that *si* is a clitic and, consequently, does not have the syntactic properties of ordinary NPs.

⁹ The generic operator *Gn* is a null universal adverb of quantification with a special modal character, which enables it to tolerate exceptions (Chierchia 1995:111).

Sentence (21a) can get an episodic reading, represented in (21b), where *Gn* is absent. This reading is prominent in contexts such as (22), where the question-answer pairs bring out the episodic readings of *si*. The truth conditions we get in (22b) are that there is a group of people, from a contextually specified set (which may or may not include the speaker or the hearer and may or may not extend to the totality of the relevant people) that is engaged in singing.

(22) a. Q: Che sta succedendo qui? A: **Si** canta. (Chierchia 1995:122)

Q: What is going on here? A: people are singing.

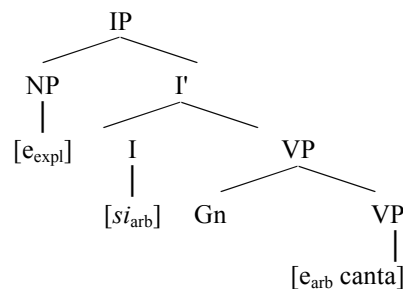
b. Q: Cosa è successi ieri in campeggio? A: **Si** è cantato.

(Chierchia 1995:122)

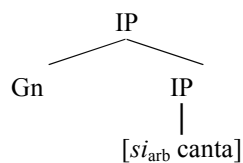
Q: What happened yesterday in the campground? A: People sang.

The sentence (21a) can also get a generic reading. This happens when there is a *Gn* operator in the LF. For (21a) there are two possible generic readings (21c) and (21d), which correspond with two different structures (24) and (23) respectively, depending on how the scope of *Gn* is selected.

(23) (Chierchia 1995:122)



(24) (Chierchia 1995:122)



In (24) the *Gn* operator is construed as having scope over IP and thus, over *si*¹⁰. Thus in the LF (21c) the *Gn* operator scopes over existential operator. This formula says that in every contextually relevant situation there is singing on. The variable C is a contextual variable that means that unless a suitable value for C is recoverable from the context, we will not be able to interpret (21c). In (23) it is *si* that has scope over the *Gn* operator

¹⁰ See footnote 6.

as represented in (21d). In this case, we first form a generic property that then gets existentially closed by *si*. This formula (21d) says that there is some groups of people (perhaps drawn from a salient set) that has the habit of singing (i.e. such that in every situation involving them where the right triggering factors are present, they sing). This "specific indefinite" interpretation is hard to get in isolation due to pragmatic reasons.

To sum up so far, the quantificational force of *si* correlates with the presence or absence at LF of the *Gn* operator or of analogous overt quantificational adverbs. *Si* behaves like an ordinary indefinite (it is existentially closed by default). I will not discuss the other properties of *si* in (11) since the focus of this review is the temporal interpretation and the arbitrary vs. generic reference of *si*.

Mendikoetxea (2002) addresses the semantic interpretation of impersonal *se* constructions in Spanish such as those in (25a), which is similar to the interpretation of sentences with a subject PRO_{arb} as in (25b). This parallelism leads her to state that both constructions have a PRO as their syntactic subject, as represented in (26) below for sentence (25a).

(25) a. Cuando **se** trabaja por placer el dinero no importa.

(Mendikoetxea 2002:236)¹¹

When **si** works for the pleasure the money not matter

"When one/people works for pleasure, the money does not matter."

b. Trabajar por placer significa que el dinero no importa.

(Mendikoetxea 2002:236)

Work_{infinitive} for the pleasure means that the money not matter

"Working for pleasure means that the money does not matter."

(26) [_{AgrP} PRO [_{Agr} *se*_{0-person}] trabaja por placer]

(Mendikoetxea 2002:237)

Mendikoetxea (1992) considers *se* to be the morphological realization of the person feature of the Agr (see chapter 4 section 4.2). The clitic *se* is not directly responsible for the arbitrary interpretation of (25)¹² but it is the presence of PRO_{arb} what gives such a reading. The proposal of Mendikoetxea (2002) can be summarized as follows:

¹¹ Glosses and translations of Mendikoetxea's (2002) examples are of my own.

¹² Contra Chierchia (1995) who argues that Italian *si* introduces a variable with a distinguished index x_{arb} , which accounts for the semantics of this construction.

- i. PRO_{arb} is the responsible of the arbitrary interpretation of the impersonal *se* constructions.
- ii. PRO_{arb} behaves like an indefinite NP, and it gets a universal or existential reading depending upon the operator (\exists or \forall) that binds it.
- iii. The interpretation of the implicit subject of this construction depends on semantic factors that follow, partially, from the type of predicate: minimal vs. non-minimal and individual vs. stage-level.

This analysis accounts for the quantificational force of the subject¹³, as well as for two facts that had so far not yet been explained: first, *se* is incompatible with existential verbs and second, *se* with minimal predicates require (most of the times) the presence of adverbials and secondary predicates.

Cinque (1988) predicts that there is no existential reading for impersonal *si/se* constructions with unaccusative verbs¹⁴. Hence, impersonal *si/se* constructions with unaccusative verbs are odd in contexts with perfective temporal reference. This can be seen in (27) in contrast with (28), where temporal reference is generic and the interpretation of the subject of the impersonal *se/si* construction is universal.

- (27) a. ***Se** entró mucho en este bar en invierno. *(existential, De Miguel 1992:157)*
 ***Se** entered a lot in this pub in the winter
 "One/people entered a lot in this pub during the winter."
- b. ***Ayer se** llegó tarde a trabajar. *(existential, De Miguel 1992:157)*
 *Yesterday **se** came late to work
 "One/people came late to work yesterday."
- c. ***Hoy se** ha nacido mucho en los hospitales madrileños.
(existential, De Miguel 1992:157)
 *Today **se** has borned a lot in the hospitals of-Madrid
- d. ***Se** murió sin dignidad en Vietnam. *(existential, De Miguel 1992:157)*
 ***Se** died without dignity in Vietnam
 "One/people died without any dignity in Vietnam."

¹³ In a way that it is not dependent on the temporal reference, contra Cinque (1988) for Italian and De Miguel (1992) for Spanish.

¹⁴ Due to thematic reasons, see Cinque (1988) and the discussion above.

- e. ***Se** fue honesto aquella tarde. (existential, De Miguel 1992:157)
 ***Se** was honest that afternoon
 "One/people was honest that afternoon."
- (28) a. **Se** entra por aquí. (universal, De Miguel 1992:157)
Se enters by here
 "One/people enters this way."
- b. **Se** llega tarde sólo cuando es inevitable. (universal, De Miguel 1992:157)
Se comes late only when is inevitable
 "One/people comes late when it is inevitable."
- c. Siempre **se** nace con poco pelo. (universal, De Miguel 1992:157)
 Always **se** is-borned with little hair
 "One/people always is borned with little hair."
- d. **Se** muere sin dignidad cuando **se** ha vivido sin amor. (universal, De Miguel 1992:157)
Se dies without dignity when **se** has lived without love
 "One/people dies without dignity when one/people has lived without love."
- e. **Se** es honrado o **se** es un trepa. (universal, De Miguel 1992:157)
Se is honest or **se** is social climber.
 "One/people is honest or social climber."

Mendikoetxea (2002) argues that the interpretation of PRO in impersonal *se* constructions is similar to the interpretation of indefinite NPs in the analysis of Diesing's (1992): PRO needs to get out of the VP so as to get the universal interpretation (29i), which is achieved by means of an operator-variable relation. The quantificational operator takes as its scope the VP and the element in [Spec, TP] as its restriction. If PRO remains within VP as in (29ii), this undergoes existential closure, the existential operator binds PRO, and the existential reading arises.

- (29) [_{AgrSP} [_{AgrS} *se*] ... [_{vP} PRO V ...]] (Mendikoetxea 2002:247)
- i. [_{AgrSP} **PRO** [_{AgrS} *se*] ... [_{vP} ~~PRO~~ V ...]] (universal; Mendikoetxea 2002:247)
- ii. [_{AgrSP} [_{AgrS} *se*] ... [_{vP} **PRO** V ...]] (existential; Mendikoetxea 2002:247)

Mendikoetxea (2002) argues contra Cinque (1988) that the fact that with non-minimal predicates¹⁵ (i.e. transitive and unergative verbs) the impersonal *se/si* construction can get a universal reading (30a) and an existential one (30b), is not related at all with the thematic properties of the verbs. Moreover, the existential reading does not depend on the specific temporal reference. It is due to the presence of a Loc (locative¹⁶) argument. See that (30b) can also be interpreted as having universal force. This is contra the prediction of Cinque's (1988).

- (30) a. Se trabaja más cuando el paro amenaza.
(universal; Mendikoetxea 2002:262)
 Se works more when the unemployment threatens
"One works more when there may be unemployment"
- b. Se trabajó mucho para levantar el país después de la guerra.
(existential; Mendikoetxea 2002:262)
 Se worked a lot to rise the country after the war
"One/people worked a lot to rise the country after the war."

Not always is there an association between the universal interpretation and the generic temporal reference either. The existential reading (31b) of (31a) is also available for impersonal *se* constructions with non-minimal predicates as the sentences in (32) and (33). Hence, the existential vs. universal readings are not related with the specific vs. generic time references but reflect whether the PRO subject of the impersonal *si* clause gets out of the VP and bound by the operator, or remains within vP and gets bound by the existential operator by existential closure.

- (31) a. Firemen are available. (Mendikoetxea 2008:324)
 b. $\forall t$ [t is time] $\exists x$ firemen(x) \wedge available(x) at t (Mendikoetxea 2008:325)
- (32) a. En estas reuniones siempre se habla de lo mismo. (Mendikoetxea 2008:325)
 In those meetings always se talk_{singular} of the same

¹⁵ Mendikoetxea (2002) argues that *minimal predicates* are 1-place predicates such as unaccusative verbs. Predicates that are 2 or more-place predicates are *non-minimal predicates*. Unergative verbs are *non-minimal predicates* because they are thought to have a (covert) object that has been conflated, along the lines proposed by Hale & Keyser (2000).

¹⁶ See Mendikoetxea (2002) for further details.

b. Generic reading

≈"In these meetings the people always talk of the same things"

c. Existential - generic reading

$\forall t$ [t is time] Loc [in these meetings] | $\exists x$ speak-of-the-same-things(x)

≈"It is always the case that in these meetings, there are people who speak of the same things."

- (33) a. Ayer se habló de política en la universidad. (Mendikoetxea 2008:325)
 Yesterday se talked_{singular} of politics in the university

b. Existential - specific reading

≈"Yesterday there were people who talked of politics in the faculty (because there was a colloquium)."

c. Universal - specific reading

≈"Yesterday, people/everybody spoke of politics in the faculty."

The quantificational force of the subject in the impersonal *si* construction does not depend on the temporal specification but it depends on how the Loc argument is interpreted: if Loc is interpreted as the subject of the predication, the sentence will be interpreted as existential. If Loc is interpreted as the restrictor of the variable, the subject of sentence will be interpreted as having universal force.

In conclusion, Mendikoetxea demonstrates that the quantificational force of the implicit subject of the impersonal *se* construction depends on the operators that bind its syntactic subject, PRO, which is interpreted as an indefinite NP. Nevertheless, it is not the case that the quantificational force necessarily correlates with the temporal reference, as Cinque (1998) argues. It depends on the position of the subject of the impersonal *se* clause (PRO) at LF: if it is out of the vP , it can be bound by an operator (e.g. universal). If PRO remains within vP , it gets bound by the existential operator by existential closure. Both options are available with both specific and generic time references. Mendikoetxea provides data that clearly show that existential readings are possible with generic time references (32c), and so are universal readings with specific time references (33c).

7.3. SE-anaphors at the syntax-semantics interface

The hypothesis that I will put forward and develop in this section is that in arbitrary constructions, a SE-anaphor is inserted in subject position, be this anaphor PRO (in

arbitrary control constructions) or PRO' and materialized in *se/si* (in Romance impersonal and passive constructions). The SE-anaphor in subject position bears Case and forms a theta-chain with the verb, so they fulfil the Double Chain Condition introduced in chapter 4, and thus it is interpreted as a verbal argument and so, as a participant of the event denoted by the verb.

7.3.1. Nominative vs. dative SE-anaphors

7.3.1.1. *ARB PRO is null PRO*

In arbitrary constructions, the subject is the null SE-anaphor PRO (as introduced in chapter 3) and the verb takes its infinitival form due to its lack of tense and Agr-features. Hence, PRO needs to be Case-marked by the matrix tense node: in (34a) PRO_{arb} is licensed by long and indirect ECM (see chapter 3) by matrix tense *I+is*, as represented in (35).

(34) a. Reading the Silmarillion is difficult.

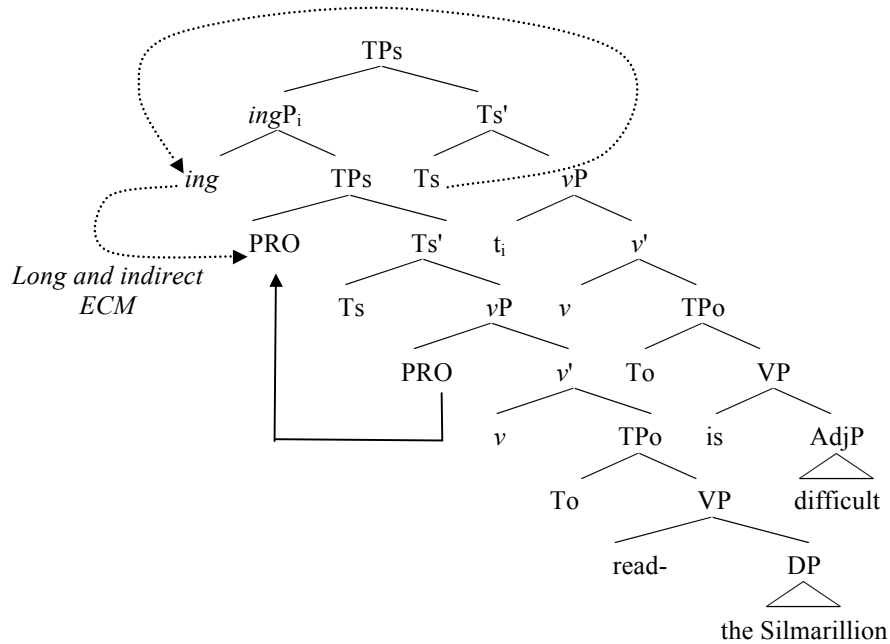
$[_{TsP} [_{ingP} PRO \text{ reading the Silmarillion }] \text{ is difficult}] \rightarrow PRO = arb$

b. Reading the Silmarillion is difficult for me.

$[_{TsP} [_{ingP} PRO \text{ reading the Silmarillion }] \text{ is difficult for me}] \rightarrow PRO = me$

Note that PRO has a grammatical number feature and hence, it is [+R]. If it can be bound by an antecedent although this is non-local, it will not take the arb value. In (34b) PRO takes the value of *me* by means of valuation at the discourse storage, since it is a suitable controller. If there is no suitable antecedent, or no antecedent whatsoever available in the discourse storage, it will get the arb value as will be explained in section 7.3.2 below.

(35) Reading the Silmarillion is difficult.



7.3.1.2. Impersonal SE is overt nominative PRO'

In this construction, exemplified in (36), the SE-anaphor that is merged in subject position is PRO' rather than PRO. Since the sentence is finite, Ts is specified for Tns-feature (Case) and PRO' is assigned nominative Case as it agrees with Ts and both share the Tns-s feature.

(36) a. Resulta difícil que se lea los libros

[_{TsP} Resulta difícil [_{CP} que se lea los libros]] → se = arb

[_{TsP} Is difficult [_{CP} that se read_{singular} the books]]

"It is difficult that the books get read."

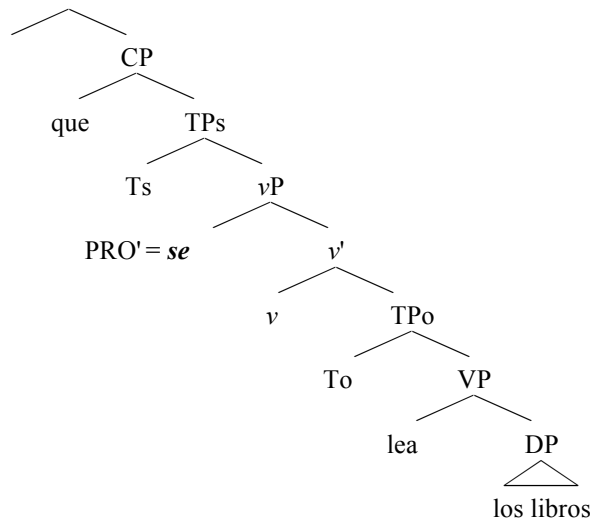
b. Me resulta difícil que se lea el Silmarillion

[_{TsP} Me resulta difícil [_{CP} que se lea los libros]] → se = arb /*me

[_{TsP} Me is difficult [_{CP} that se read_{singular} the books]]

"For me it is difficult that the books get read."

- (37) ... que **se** lea los libros.
 ... that **se** read_{singular} the books_{plural}
 "... that the books get read."



The number φ -feature of T_S remain unvalued (although it is interpretable on PRO') and the person φ -feature of T_S gets valued by agreeing with PRO'. These φ -features on T_S (which end up on the verb in PF) are materialized as third person and default number, which varies from language to language: in Spanish it is singular, whereas in Italian it is plural.

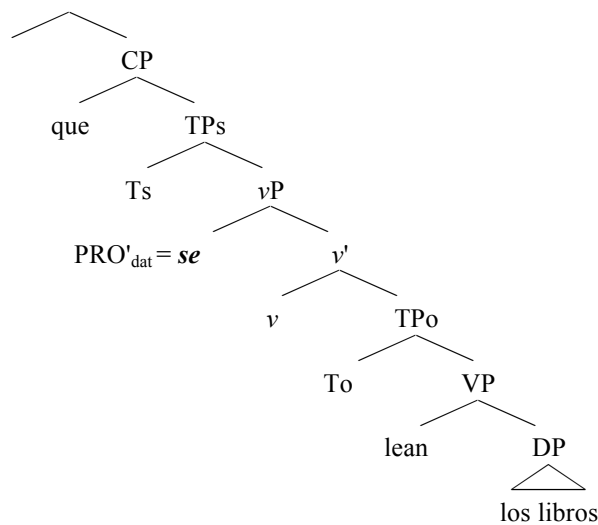
7.3.1.3. Passive SE is overt dative PRO'

In this construction, as in (38) below, the SE-anaphor PRO' is merged in subject position, as in impersonal constructions. The difference is that the PRO' that is merged here is different from that of the impersonal constructions: this PRO' is marked in the lexicon with inherent dative Case.

- (38) a. Resulta difícil que **se** lean los libros
 [TP_S Resulta difícil [CP que **se**_{dat} lean los libros]] → **se** = arb
 [TP_S Is difficult [CP that **se**_{dat} read_{plural} the books]]
 "It is difficult that the books get read."
 b. **Me** resulta difícil que **se** lean los libros
 [TP_S **Me** resulta difícil [CP que **se**_{dat} lean los libros]] → **se** = arb /*me
 [TP_S **Me** is difficult [CP that **se**_{dat} read_{plural} the books]]
 "For me it is difficult that the books get read."

Since *se* has inherent dative Case, it does not need to form a chain with Ts. It does form a chain with the verb so that it values its theta-feature and hence, it is interpreted as a verbal argument. Since the φ -features of Ts are unvalued and uninterpretable, Ts searches in its c-command domain and agrees with the DP object via vP , and hence the number feature on Ts agrees with the number feature on the DP object.

- (39) ... que **se** lean los libros.
 ... that **se** read_{plural} the books_{plural}
 "... that the books get read."



It can be seen in (38b) that impersonal *se/si* cannot get the value of *me* (me) by means of (co-)valuation due to its being [-R].

7.3.1.4. Arbitrary SE in Germanic

In Germanic languages, in particular in Dutch and English, arbitrary SE is impossible because in English, PRO' is never realized and PRO is incompatible with tensed clauses, since Ts would value its φ -features. In Dutch, PRO' cannot be merged in subject position because if PRO' is nominative-marked it would not be pronounced, as seen in chapters 5 and 6.

7.3.2. Arbitrary interpretation and choice functions

I have reviewed the most relevant theories on the interpretation of arbitrary constructions. We saw that both Chierchia (1995) and Mendikoetxea (2002) argue that impersonal *se/si* is interpreted as an indefinite NP. Mendikoetxea (2008) resorts to a generic operator in subject position. The generic operator introduces a variable that has

to be bound by an operator, and this variable ranges between the existential and universal readings. The variability of range of the variable introduced by the generic operator is best captured by choice functions, which were used by Reinhart (1997, 2006) to account for certain interpretations of weak indefinites.

Reinhart (1997, 2006) introduces *choice functions* in the scope of quantifiers because the actual options of covert scope appear inconsistent with what would be entailed by one well-behaved syntactic operation like *Quantifier Raising* (QR). There are some indefinite NPs that do not behave as standard *generalized quantifiers* (GQ) over singular individuals but they lack a (CG) determiner, and thus, they are locally interpreted by choice functions. The interpretative problem that Reinhart faces is how to assign wide scope to existential NPs that, otherwise, show properties of remaining *in situ*. Her solution to this problem is to allow existential quantification over choice functions as defined in (40).

(40) Choice function: (Reinhart 2006:81)

A function f is a choice function (CH(f)) if it applies to any non-empty set and yields a member of that set.

The fact that a choice function cannot apply to an empty set was observed by Zermelo (1904) and formalized in his *Axiom of Choice*. Note that the generic operator can never give as output the zero value either.

Since we already have in the semantic machinery a device that is able to range from existential to universal while disallowing the zero value, it is more economic to use such a device (the choice function) rather than make use of a generic operator that gives the same results but requires more restrictions to be empirically correct. Moreover, the use of choice functions solves the problems posed by *Quantifier Raising* and quantifiers interpreted *in situ* in some of contexts (see Reinhart 1997 for more details on the use of choice functions vs. *Quantifier Raising*).

The hypothesis that I put forward is that PRO', being [-R], can only be interpreted (if it is interpreted, i.e. if it fulfils the Double Chain Condition) in two ways: by A-binding (either by syntactic-chain formation in C_{HL} or directly by A-binding at C-I) or by means of a choice function. Since it is [-R], it can never get a value directly from the discourse storage. PRO_{arb}, on the other hand, it is [+R] and can get a value from the discourse, hence a suitable antecedent can cancel the arbitrary reading with PRO_{arb} (41) but not with impersonal or passive *se/si* (42).

- (41) a. Leer el Silmarillion resulta difícil
 [TPs [CP PRO leer el Silmarillion] resulta difícil] → PRO = arb
 [TPs [CP PRO read_{infinitive} the Silmarillion] is difficult]
"Reading the Silmarillion is difficult."
- b. Leer el Silmarillion **me** resulta difícil
 [TPs [CP PRO leer el Silmarillion] **me** resulta difícil] → PRO = *arb/me
 [TPs [CP PRO read_{infinitive} the Silmarillion] **me** is difficult]
"Reading the Silmarillion is difficult for me."
- (42) a. Resulta difícil que **se** lea el Silmarillion
 [TPs Resulta difícil [CP que **se** lea el Silmarillion]] → **se** = arb
 [TPs Is difficult [CP that **se** read the Silmarillion]
"It is difficult that the Silmarillion gets read."
- b. **Me** resulta difícil que **se** lea el Silmarillion
 [TPs **Me** resulta difícil [CP que **se** lea el Silmarillion]] → **se** = arb/*me
 [TPs **Me** is difficult [CP that **se** read the Silmarillion]
"It is difficult for me that the Silmarillion gets read."

How are sentences (41) and (42) interpreted? Both PRO and PRO' are interpreted as weak indefinites by means of a choice function that ranges in a group which is unrestricted, so it is universal. This means that the choice function can give a value from 1 to infinite, but never a 0 due to the *Axiom of Choice*. The interpretation of (41a) and (42a,b) is represented in (43):

$$(43) \exists f (CH(f) \wedge \forall z \rightarrow f(z) \text{ read (The Silmarillion)})$$

Formula (43) can be roughly paraphrased as "there exists a function f such that f is a choice function and there exists a set composed of the universe of z such that the function f applies to the set of z and gives a value (from 1 to universal) such that that value reads the Silmarillion". That *se/si* and PRO_{arb} is interpreted as being human is just a pragmatic epiphenomenon, such as Mendikoetxea (2002) shows.

To sum up, I have introduced an analysis of arbitrary constructions as being derived by the presence of SE-anaphors that are interpreted as weak indefinites, i.e. by means of choice functions that are able to select from zero to infinite individuals, hence the generic interpretation and the variation between the quasi-existential and quasi-universal quantification of the subject of arbitrary constructions.

7.4. Conclusions

In summary, we have seen in this chapter a recapitulation of the environments where the SE-anaphors appear and I have proposed a still tentative analysis of arbitrary environments with SE-anaphors. The picture of what has been proposed in this chapter and the previous ones is summarized in table (44) below:

(44) Constructions with SE-anaphors:

	English	Dutch	Romance	Slavic
OC	PRO \Rightarrow \emptyset	PRO \Rightarrow \emptyset	PRO \Rightarrow \emptyset	PRO \Rightarrow \emptyset
NOC	PRO \Rightarrow \emptyset	PRO \Rightarrow \emptyset	PRO \Rightarrow \emptyset	PRO \Rightarrow \emptyset
Reflexive	PRO' \Rightarrow \emptyset	PRO' \Rightarrow zich	PRO' \Rightarrow se/si	PRO' \Rightarrow się
Ergative	PRO' \Rightarrow \emptyset	PRO' \Rightarrow \emptyset	PRO' \Rightarrow se/si	PRO' \Rightarrow się
Inherent	PRO' \Rightarrow \emptyset	PRO' \Rightarrow zich	PRO' \Rightarrow se/si	PRO' \Rightarrow się
Aspectual	verbal particles	verbal particles	PRO' \Rightarrow se/si	?
AC	PRO \Rightarrow \emptyset	PRO \Rightarrow \emptyset	PRO \Rightarrow \emptyset	PRO \Rightarrow \emptyset
Impersonal	does no exist	does not exist	PRO' \Rightarrow se/si	PRO' \Rightarrow się
Passive	does not exist	does not exist	PRO' _{DAT} \Rightarrow se/si	PRO' _{DAT} \Rightarrow się

Leaving aside the middle constructions, we see that a series of what seemed unrelated constructions, can be accounted for by a unified analysis that rests on the presence of SE-anaphors and their processing at the linguistic interfaces.

Chapter 8

Conclusions

In this dissertation we have looked at several kinds of structures: first, control structures, both obligatory (1a) and non-obligatory (1b). Second, reflexive verbs, both inherent (2a,c) and non-inherent (2b). Third, other verbs that require the presence of what seems to be a reflexive marker, such as theme unaccusative verbs in (3a) and subject experiencer verbs in (3b). Finally, I have addressed arbitrary constructions as arbitrary control in (4), impersonal *se/si* in (5) and passive *se/si* in (5).

- (1) a. John_i wants [**PRO**_{i/*j} to read The Silmarillion].
b. [**PRO**_{i/j} Reading the Silmarillion] is difficult for John_i.
- (2) a. John_i washed (**himself**_{i/*j}).
b. John_i hates *(**himself**_{i/*j}).
c. Jan_i waste **zich** (**zichzelf**_{i/*j})
Jan_i washed **zich** (**zichzelf**_{i/*j})
"Jan washed (himself)."
- (3) a. El cristal *(**se**) rompió. *(Mendikoetxea 2008:291)*
The glass *(**se**) broke
"The glass broke."
b. Juan *(**se**) sorprendió.
Juan *(**se**) got-surprised
"Juan got surprised."
- (4) [**PRO**_{i/j} Reading the Silmarillion] is difficult.
- (5) **Se** come las manzanas. *(Mendikoetxea 2008:304)*
Se eat_{singular} the apples_{plural}
"People / one eats the apples."
- (6) **Se** comen las manzanas. *(Mendikoetxea 2008:291)*
Se eat_{plural} the apples_{plural}
"The apples are eaten."

My claim is that all these structures (1) to (6) have in common the presence of a SE-anaphor. This kind of anaphors is usually deficient in referentiality (though not in the case of non-obligatory controlled PRO as in (1b)) due their lack of grammatical number, also deficient in other φ -features such as gender (though not *pro* which has a full set of φ -features), and many times phonologically defective (they are usually either null or clitics). Finally, they allow local binding iff the theta-grid of the verb has somehow been altered. The arbitrary interpretation of the SE-anaphors is reached by means of an existential closure of a choice function rather than by the presence of a generic operator. In the cases of *pro* I have argued that another SE-anaphor exists in the lexicon: PRO*, which is similar to PRO but unlike it, PRO* has a gender φ -feature (as well as person and grammatical number). Hence, it is [+R] and when its φ -features get valued by means of agree-chains with Ts, PRO* either gets phonologically realized and end up as a pronominal or is not pronounced so that it ends up as *pro*. Non-null languages like English are assume either to lack PRO* in the lexicon or not to allow PRO* to be nominative assigned (since Rizzi (1986) showed that *pro* in object position is allowed under certain conditions).

Note that in Romance and Dutch, the inherent reflexive verbs such as (3a) require the presence of a clitic *si/se* or a particle *zich*, respectively. I claim that the abstract structures of (1) to (6) are as follows:

- (7) DP wants [SE to read DP] (control structures)
 where SE = PRO
 at PF SE \Rightarrow PRO (both in Romance and Germanic)
- (8) DP wash SE (inherent reflexive verbs)
 where SE = PRO'
 at PF SE \Rightarrow *se/si* (in Romance)
 SE \Rightarrow *zich* (in Dutch)
 SE \Rightarrow \emptyset (in English)
- (9) a. DP broke SE (theme unaccusative verbs)
 where SE = PRO'
 at PF SE \Rightarrow *se/si* (in Romance)
 SE \Rightarrow \emptyset (in Dutch)
 SE \Rightarrow \emptyset (in English)

- b. DP got-surprised SE *(subject experiencer verbs)*
- | | | |
|-------|-------------------------------|---------------------|
| where | SE = PRO' | |
| at PF | SE \Rightarrow <i>se/si</i> | <i>(in Romance)</i> |
| | SE \Rightarrow <i>zich</i> | <i>(in Dutch)</i> |
| | SE \Rightarrow \emptyset | <i>(in English)</i> |
- (10) [SE reading DP] is AdjP *(arbitrary control)*
- | | | |
|-------|------------------------------|---------------------------------------|
| where | SE=PRO | <i>(arb by CH(f))</i> |
| at PF | SE \Rightarrow \emptyset | <i>(both in Romance and Germanic)</i> |
- (11) SE eat_{no agreement with DP} DP *(impersonal se/si)*
- | | | |
|-------|-------------------------------|----------------------------------|
| where | SE = PRO' | <i>(arb by CH(f) in Romance)</i> |
| | SE = impossible | <i>(in Dutch and English)</i> |
| at PF | SE \Rightarrow <i>se/si</i> | <i>(in Romance)</i> |
- (12) SE eat_{agreement with DP} DP *(passive se/si)*
- | | | |
|-------|-------------------------------|----------------------------------|
| where | SE = PRO' | <i>(arb by CH(f) in Romance)</i> |
| | SE = impossible | <i>(in Dutch and English)</i> |
| at PF | SE \Rightarrow <i>se/si</i> | <i>(in Romance)</i> |

Whereas the morphological realization of SE (which stands for SE-anaphor) is PRO in control structures across languages (7), there are some differences in the morphological realization with reflexive (8), unaccusative (9a) and experiencer verbs (9b).

More concretely, SE is materialized as a clitic with theme unaccusative verbs in Romance but is not spelled out in Germanic (9a). On the hand, SE is materialized with subject experiencer verbs in Romance and Dutch but not in English (9b) (although SE is still present, as it happens with PRO). As for inherent reflexive verbs, both Spanish and Dutch spell out SE unlike English, which never spells out SE (8).

We have seen two different SE-anaphors: PRO, PRO'; and one null pronominal PRO*.

- A. PRO is a null SE-anaphor with a set of interpretable and unvalued φ -features (number and person), and a tense feature (Case). Since it has a grammatical number φ -feature, it can be [+R]. Depending on the syntactic structure where it appears, obligatory or non-obligatory control arises. In other words, if a compound syntactic chain can be formed between PRO and its antecedent, there is obligatory control. If no syntactic chain can be formed, PRO has to either be

bound at the C-I system or to get a value directly from the discourse storage (recall it is [+R]).

- B. PRO' is a null SE-anaphor with a defective set of interpretable and unvalued φ -features (just person), and a tense feature (Case). Since it does not have no grammatical number φ -feature, it is [-R]. It is inserted when the thematic grid of the verb has been reduced, in order to formally check or make deleteable some formal features (theta-features) on the tempo-aspectual heads of the sentence. Since it establishes agree-relations with the tempo-aspectual and verbal system, its person feature gets valued with the person feature of the subject and hence, it can, in principle, get spelled-out. This depends on some spell-out rules that hold at the syntax-phonology interface: Spanish always spells out the φ -person feature of PRO' (*se*), English never does so and Dutch sometimes (*zich*). If PRO' is inserted in subject position and fulfils the Double Chain Condition (DCC), then it is interpreted as an argument. In this case, as in the case of unbounded PRO (i.e. PRO_{arb}) the quantificational interpretation of the SE-anaphor is done by means of an existential closure on a choice function introduced by the SE-anaphor, lacking this a Generalized Quantifier determiner (like weak indefinites). Also we have seen that there exists in the lexicon an instance of PRO' with inherent Case, which I have called PRO'_{DAT}.
- C. PRO* is a null pronominal with a full set of φ -features which renders it [+R] and it can either be pronounced (and it ends up as an overt pronominal) or not (so that it ends up as *pro*).

In table (13) below the main properties of the three null pronouns available in UG:

(13) Null pronouns in UG (SE-anaphors and pronominals):

	SE-anaphors			Pronominals
	PRO	PRO'	PRO' _{dat}	PRO*
φ-features	<i>i</i> φ person [unval] <i>i</i> φ number [unval]	<i>i</i> φ person [unval]	<i>i</i> φ person [unval]	<i>i</i> φ person [unval] <i>i</i> φ number [unval] <i>i</i> φ gender [unval]
θ-features	<i>i</i> θ [unval]	<i>i</i> θ [unval]	<i>i</i> θ [unval]	<i>i</i> θ [unval]
Tns (Case)	<i>u</i> Tns [unval]	<i>u</i> Tns [unval]	<i>i</i> Tns [dative]	<i>u</i> Tns [unval]
[±R]	[±R]	[-R]	[-R]	[+R]
Long ECM	YES	YES	NO	YES
Chain formation	YES	YES	NO	NO

In table (14) we can see the different constructions studied in this dissertation and which SE-anaphor is present and triggers the different syntactic characteristics and semantic interpretation particular to the construction.

(14) Constructions with SE-anaphors studied in the dissertation:

	English	Dutch	Romance	Slavic
OC	PRO ⇒ ∅	PRO ⇒ ∅	PRO ⇒ ∅	PRO ⇒ ∅
NOC	PRO ⇒ ∅	PRO ⇒ ∅	PRO ⇒ ∅	PRO ⇒ ∅
Reflexive	PRO' ⇒ ∅	PRO' ⇒ zich	PRO' ⇒ se/si	PRO' ⇒ się
Ergative	PRO' ⇒ ∅	PRO' ⇒ ∅	PRO' ⇒ se/si	PRO' ⇒ się
Inherent	PRO' ⇒ ∅	PRO' ⇒ zich	PRO' ⇒ se/si	PRO' ⇒ się
AC	PRO ⇒ ∅	PRO ⇒ ∅	PRO ⇒ ∅	PRO ⇒ ∅
Impersonal	does not exist	does not exist	PRO' ⇒ se/si	PRO' ⇒ się
Passive	does not exist	does not exist	PRO' _{DAT} ⇒ se/si	PRO' _{DAT} ⇒ się

In table (15) below we can see a complete catalog of the null pronouns and how they are instantiated in the different syntactic constructions studied in this dissertation.

The variation that we observe in (7) to (12) across languages, Romance (Spanish) and Germanic (English and Dutch) in our case, can be accounted for by resorting to processes that take place at the lexicon-syntax (arity operations) and syntax-phonology (spell-out rules) interfaces. Hence, we pursue the issue of a universal syntactic component (Chomsky 2001, 2005, 2006; Reinhart 2002; among many others). The impersonal and passives *se/si* constructions are impossible in the Germanic languages so far studied is because in both English and Dutch, SE-anaphors cannot appear in Case position due to their impossibility of being nominative marked, or to the inexistence of a PRO' marked with inherent dative in the lexicon.

In Burzio's (1986) terms, we have studied the clitic *si*, i.e. that which appears with unaccusative and reflexive verbs, as well as SI, which appears with impersonal *se/si* constructions and, as I have argued, with arbitrary control. Middle *si* has been left as an open question for further research nonetheless. It is my belief, however, that future research will cast some light on this issue, and will improve the integration of both SI and *si* in the unified analysis I have developed in this dissertation.

(15) Instantiation of null pronouns:

	OC PRO	NOC PRO	AC PRO	reflexive <i>se/si/zich</i>	ergative <i>se/si</i>	inherent <i>se/si/zich</i>	impersonal <i>se/si</i>	passive <i>se/se</i>	<i>pro</i>	pronominal
pronoun	PRO	PRO	PRO	PRO'	PRO'	PRO'	PRO'	PRO' _{DAT}	PRO*	PRO*
materialization	∅	∅	∅	<i>se/si/zich</i>	<i>se/si</i>	<i>se/si/zich</i>	<i>se/si</i>	<i>se/si</i>	∅	he, she, you, etc.
φ-features	<i>i</i> φperson [v] <i>i</i> φnumber [v]	<i>i</i> φperson [-] <i>i</i> φnumber [-]	<i>i</i> φperson [-] <i>i</i> φnumber [-]	<i>i</i> φ person [v]	<i>i</i> φ person [v]	<i>i</i> φ person [v]	<i>i</i> φ person [-]	<i>i</i> φ person [-]	<i>i</i> φ person [v] <i>i</i> φ number [v] <i>i</i> φ gender [v]	<i>i</i> φ person [v] <i>i</i> φ number [v] <i>i</i> φ gender [v]
θ-features	<i>i</i> θ [v]	<i>i</i> θ [v]	<i>i</i> θ [v]	<i>i</i> θ [-]	<i>i</i> θ [-]	<i>i</i> θ [-]	<i>i</i> θ [v]	<i>i</i> θ [v]	<i>i</i> θ [v]	<i>i</i> θ [v]
Case	<i>u</i> Tns [ACC]	<i>u</i> Tns [ACC]	<i>u</i> Tns [ACC/NOM]	<i>u</i> Tns [ACC]	<i>u</i> Tns [NOM]	<i>u</i> Tns [ACC]	<i>u</i> Tns [NOM]	<i>i</i> Tns [DAT]	<i>u</i> Tns [NOM]	<i>u</i> Tns [NOM]
[±R]	[-R]	[+R]	[-R]	[-R]	[-R]	[-R]	[-R]	[-R]	[+R]	[+R]
DCC	YES	YES	YES	NO	NO	NO	YES	YES	YES	YES
Long ECM	YES	YES	YES	YES	YES	YES	YES	NO	YES	NO
Interpretation	A-binding by syntactic chains	A-binding at C-I or valuation	A-binding at <i>arb</i> by CH(f)	not interpreted	not interpreted	not interpreted	<i>arb</i> by CH(f)	<i>arb</i> by CH(f)	A-binding at C-I or valuation	A-binding at C-I or valuation

References

- Ackema, Peter, and Maaike Schoorlemmer. 2006. Middles. In *The Blackwell Companion to Syntax*, ed. Martin Everaert and Henk van Riemsdijk, volume III, chapter 42, 131–203. Blackwell.
- Alexiadou, Artemis, and Elena Anagnostopoulou. 1998. Parametrizing AGR: Word order, V-movement and EPP-checking. *Natural Language and Linguistic Theory* 16:491–539.
- Andrews, Avery D. 1971. Case agreement of predicate modifiers in Ancient Greek. *Linguistic Inquiry* 2:127–151.
- Andrews, Avery D. 1976. The VP-complement analysis in modern Icelandic. In *Proceedings of NELS*, volume 6, 1–21.
- Aranovich, Raúl. 2000. Split intransitivity and reflexives in Spanish. *Probus* 12:165–186.
- Belletti, A. 1982. "Morphological passive" and *pro*-drop: The impersonal construction in Italian. *Journal of Linguistic Research* 2.4:1–34.
- Belletti, Adriana. 1987. Los inacusativos como asignadores de caso. In *Sintaxis de las Lenguas Románicas*, ed. Violeta Demonte and Marina Fernández Lagunilla. Madrid: El Arquero.
- Bello, Andrés. 1841. *Análisis Ideológica de los Tiempos de la Conjugación Castellana*. Calparaíso, Imprenta de M Rivadeneyra.
- Boeckx, Cedric, and Norbert Hornstein. 2004. Movement under control. *Linguistic Inquiry* 35:431–452.
- Borer, Hagit. 1985. *Parametric syntax*. Dordrecht: Foris.
- Bouchard, Denis. 1984. *On the Content of Empty Categories*. Dordrecht: Foris.
- Burzio, Luigi. 1986. *Italian Syntax. A Government-Binding Approach*. Dordrecht: Kluwer.
- Cardinaletti, Anna. 2004. Toward cartography of subject positions. In *The structure of CP and IP*, ed. L. Rizzi. Oxford: Oxford University Press.
- Cecchetto, Carlo, and Renato Oniga. 2001. Consequences of the analysis of Latin infinitival clauses for the theory of case and control. Ms. University of Milano-Bicocca and University of Udine.

- Cecchetto, Carlo, and Renato Oniga. 2004. A challenge to null case theory. *Linguistic Inquiry* 141–149.
- Chierchia, Gennaro. 1989. A semantics for unaccusatives and its syntactic consequences. Ms. Cornell University.
- Chierchia, Gennaro. 1995. The variability of impersonal subjects. In *Quantification in Natural Languages*, ed. E. Bach, E. Jelinek, A. Kratzer, and B. H. Partee, 107–143. Kluwer Academic Publishers.
- Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
- Chomsky, Noam. 1982. Some concepts and consequences of the theory of government and binding. *Linguist Inquiry Monograph Six*.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge MA: The MIT Press.
- Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: A Life in Language*, ed. M. Kenstowicz, 1–52. MA: The MIT Press.
- Chomsky, Noam. 2005. Three factors in language design. *Linguistic Inquiry* 36:1–22.
- Chomsky, Noam. 2006. On phases. In *Foundational Issues in Linguistic Theory*, ed. Robert Freidin, Carlos Peregrin Otero, and Maria Luisa Zubizarreta. Cambridge MA: MIT Press. MIT.
- Chomsky, Noam, and Howard Lasnik. 1993. The theory of Principles and Parameters. In *Syntax: An International Handbook of Contemporary Research*, ed. Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld, and Theo Vennemann, 506–569. Berlin: Mouton de Gruyter.
- Cinque, Guglielmo. 1988. On *si* constructions and the theory of *arb*. *Linguistic Inquiry* 19:521–581.
- Cinque, Guglielmo. 1999. *Adverbs and Functional Heads: A Cross-linguistic Perspective*. New York: Oxford University Press.
- Comorovski, Ileana. 1985. Control and obviation in Romanian. In *Proceedings of ESCOL*, ed. Choi Soonja, Dan Devitt, Wynn Janis, Terry McCoy, and Zhengsheng Zhang, 47–56. Ohio State University.
- Culicover, Peter W., and Ray Jackendoff. 2001. Control is not movement. *Linguistic Inquiry* 32:493–512.
- De Miguel, Elena. 1992. *El Aspecto en la Sintaxis del Español*. Madrid: Ediciones de la Universidad Autónoma de Madrid.
- De Miguel, Elena, and Marina Fernandez Lagunilla. 2000. El operador aspectual "se". *Revista Española de Lingüística* 30:13–44.

- Diesing, Molly. 1992. Indefinites. *Linguist Inquiry Monograph* 20.
- Dinnsen, D. 1972. Additional constraints on clitic order in Spanish. In *Generative studies in romance languages*, ed. J. Casagrande and S. Saciuk. Rowley Mass: Newbury House.
- Dobrovie-Sorin, Carmen. 1998. Impersonal *se* constructions in Romance and the passivization of unergatives. *Linguistic Inquiry* 29:399–437.
- Dobrovie-Sorin, Carmen. 2006. The *SE*-anaphor and its role in argument realization. In *The Blackwell Companion to Syntax*, ed. Martin Everaert and Henk van Riemsdijk, volume IV, chapter 56. Blackwell.
- Doron, Edit, and Malka Rappaport-Hovav. 2007. On the uniform application of valence-changing operations. In *IATL 23*. Tel-Aviv University.
- Epstein, Samuel D., Hisatsugu Kitahara, and T. Daniel Seely. 2008. Uninterpretable features: What are they and what do they do? Ms. June 2008.
- Everaert, Martin. 1986. *The Syntax of Reflexivization*. Dordrecht: Foris.
- Everaert, Martin. 2002. The case of the Theta System. *Theoretical Linguistics* 28.
- Fernandez Lagunilla, Marina. 1987. Los infinitivos con sujeto léxico en español. In *Sintaxis de las lenguas románicas*, ed. Violeta Demonte and Marina Fernández Lagunilla. Madrid: El Arquero.
- Fernández Soriano, Olga. 1989. Strong pronouns in null subject languages and the avoid pronoun principle. *MIT Working Papers in Linguistics* 11:228–240.
- Fernández Soriano, Olga. 1993. *Los Pronombres Átonos*. Madrid: Taurus Universitaria.
- Fernández Soriano, Olga. 1999. Datives in constructions with unaccusative *se*. *Catalan Working Papers in Linguistics* 7.
- Fillmore, Charles J. 1971. *Lectures on Deixis*. CSLI Publications.
- Gallardo, Elías. 2008. Estrategias de inacusatividad en inglés, japonés y lenguas románicas. In *Actas del VIII Congreso de Lingüística General*, ed. Antonio Moreno Sandoval. Universidad Autónoma de Madrid, Madrid: Universidad Autónoma de Madrid.
- González Vergara, Carlos. 2006. *Las Construcciones no Reflexivas con "se". Una Propuesta desde la Gramática del Papel y la Referencia*. Doctoral Dissertation, Universidad Complutense de Madrid e Instituto Universitario de Investigación Ortega y Gasset, Madrid.
- Grimshaw, Jane. 1990. *Argument Structure*. Cambridge MA: MIT Press.
- Grimshaw, Jane. 1991. *Extended Projection*. Ms. Brandeis University.

- Grimshaw, Jane. 2000. Locality and extended projection. In *Lexical Specification and Insertion*, ed. Peter Coopmans, Martin Everaert, and Jane Grimshaw. Amsterdam: John Benjamins.
- Hale, Ken, & Hay Keyser. 1997. The limits of argument structure. In *Theoretical Issues at the Morphology-Syntax Interface*, ed. Amaya Mendikoetxea and Myriam Uribe-Etxebarria. Biskaia: Servicio Editorial de la UPV.
- Hale, Ken, and Jay Kayser. 2000. Conflation. In *IUIOG Working Papers in Linguistics*, ed. Ana Bravo Martín, Carlos Luján Berenguel, and Isabel Pérez Jiménez, volume 7, 39–76. Madrid: Instituto Universitario de Investigación Ortega y Gasset.
- Halliday, M., and R. Hasan. 1989. *Language, Context, and Text: Aspects on Language in a Social-Semiotic Perspective*. Oxford: Oxford University Press.
- Harley, Heidi. 2000. Irish, the EPP and PRO. University of Arizona.
- Heim, Irene. 1982. *The Semantics of Definite and Indefinite NPs*. Doctoral Dissertation, MIT.
- Hernanz, Maria Luisa. 1994. Argumentos implícitos, operadores nulos e interpretación arbitraria. El caso de los infinitivos pseudoecuativos. *Gramática del Español* 6:315–362.
- Hernanz, Maria Luisa. 1999. El infinitivo. In *Gramática Descriptiva de la Lengua Española*, ed. Ignacio Bosque and Violeta Demonte, volume 2, chapter 36. Madrid: Espasa Calpe.
- Holmberg, Anders. 2005. Is there a little *pro*? Evidence from Finnish. *Linguistic Inquiry* 36:533–564.
- Horn, Laurence R. 2008. "I love me some him": The landscape of non-argument datives. *Empirical Issues in Syntax and Semantics* 7:169–192.
- Hornstein, Norbert. 1999. Movement and control. *Linguistic Inquiry* 30.1:69–96.
- Hornstein, Norbert. 2003. On control. In *Minimalist syntax*, ed. Randall Hendrick. Blackwell.
- Horvath, Julia, and Tal Siloni. 2008. Causatives across components. Tel Aviv University.
- Jackendoff, Ray. 1987. The status of thematic relations in linguistic theory. *Linguistic Inquiry* 18:369–411.
- Kayne, Richard S. 1975. *French Syntax: The Transformational Cycle*. Cambridge MA: MIT Press.

- Kayne, Richard S. 1984. *Connectedness and Binary Branching*. Dordrecht: Kluwer.
- Kayne, Richard S. 1988. Romance *se/si*. *GLOW Newsletter* 20.
- Kayne, Richard S. 1991. Romance clitics, verb movement and PRO. *Linguist Inquiry* 22:647–686.
- Kayne, Richard S. 2000. *Parameters and Universals*. New York: Oxford University Press.
- Kayne, Richard S. 2008. Comparative syntax and the lexicon. Ms. New York University, October 2008.
- Kersten, Johan, Eddy Ruys, and Joost Zwarts. 2001. Lexicon of linguistics. <http://www2.let.uu.nl/UiL-OTS/Lexicon/>.
- Labelle, Marie. 2008. The French reflexive and reciprocal *se*. *Natural Language and Linguistic Theory* 26:833–876.
- Landau, Idan. 2003. Movement out of control. *Linguistic Inquiry* 34:471–498.
- Landau, Idan. 2004a. Movement-resistant aspects of control. In *New Horizons in the Grammar of Raising and Control*, ed. Stanley Dubinsky and William Davies. Springer.
- Landau, Idan. 2004b. The scale of finiteness and the Calculus of Control. *Natural Language and Linguistic Theory* 22:811–877.
- Landau, Idan. 2006. Severing the distribution of PRO from case. *Syntax* 9:153–170.
- Langacker, R. 1991. *Foundations of Cognitive Grammar*, volume 2: Descriptive Application. Stanford: Stanford University Press.
- Lebeaux, David. 2009. *Where Does Binding Theory Apply?* Linguistic Inquiry Monograph Fifty. Cambridge MA: MIT Press.
- Levin, Beth, and Malka Rappaport. 1995. *Unaccusativity at the Syntax-Lexical Semantics Interface*. Cambridge MA: MIT Press.
- Lozano, Cristóbal. 2008. *The acquisition of syntax and discourse: Pronominals and word order in English and Greek learners of Spanish*. Saarbrücken: VDM Verlag.
- Lozano, Cristóbal, & Amaya Mendikoetxea. 2009. L2 syntax meets information structure: Word order at the interfaces. In *BUCLD 33: Proceedings of the 33rd Annual Boston University Conference on Language Development*, ed. J. Chandlee, M. Franchini, S. Lord, and M. Rheiner, volume 1, 313–324. Somerville (Mass): Cascadilla Press.

- Lozano, Cristóbal, & Amaya Mendikoetxea. forthcoming. Interface conditions on postverbal subjects: A corpus study of L2 English. *Bilingualism: Language and Cognition*.
- Luján, Marta. 1999. A unified approach to control and obviation.
- Maldonado, Ricardo. 1999. *A Media Voz. Problemas Conceptuales del Clítico "se"*. México D.F.: Universidad Nacional Autónoma de México.
- Manzini, M. Rita. 1983. On control and control theory. *Linguistic Inquiry* 14:421–446.
- Manzini, M. Rita, and Anna Roussou. 2000. A minimalist theory of A-movement and control. *Lingua* 110:409–447.
- Manzini, Maria Rita. 1986. On italian *si*. In *Syntax and semantics 19: The syntax of pronominal clitics*, ed. Hagit Borer, 241–262. New York: Academic Press.
- Marantz, A. 1984. *On the Nature of Grammatical Relations*. Cambridge MA: MIT Press.
- Marelj, Marijana. 2004. *Middels and Argument Structure across Languages*. Doctoral Dissertation, University Utrecht.
- Martin Zorraquiño, M.A. 1979. *Las Construcciones Pronominales en Español*. Madrid: Gredos.
- Masullo, Pascual José. 1999. *La Interfaz Léxico-Sintaxis: Presencia y Ausencia del Clítico se en Construcciones Inacusativas*. Ms. Universidad Nacional del Comahue y University of Washington, march 1999.
- Mendikoetxea, Amaya. 1992. *On the Nature of Agreement: The Syntax of arb se in Spanish*. Doctoral Dissertation, York University.
- Mendikoetxea, Amaya. 1997. Clitics as Agr, and PRO in finite clauses. In *Proceedings o the IV Colloquium of Generative Grammar*, ed. Natalia Catalá and María Bargalló.
- Mendikoetxea, Amaya. 1999. Construcciones con *se*. Medias, pasivas e impersonales. In *Gramática Descriptiva de la Lengua Española*, ed. Ignacio Bosque and Violeta Demonte, chapter 26. Madrid: Espasa-Calpe.
- Mendikoetxea, Amaya. 2000. Relaciones de interficie: Los verbos de cambio de estado. In *Cuadernos de Lingüística VII*, ed. A. Bravo, C. Lujan, and I. Pérez, 125–144. Instituto Universitario Ortega y Gasset.
- Mendikoetxea, Amaya. 2002. La semántica de la impersonalidad. In *Las Construcciones con se*, ed. C. Sanchez Lopez, 239–267. Madrid: Visor.

- Mendikoetxea, Amaya. 2008. Clitic impersonal constructions in romance: Syntactic features and semantic interpretation. In *Transactions of the Philological Society*, ed. Paul Rowlett and Anna Siewierska, volume 106, 290–336. Oxford: Blackwell.
- Nevins, Andrew. 2004. Derivations without the Activity Condition. In *Perspective on Phases. MITWPS*, ed. Martha McGinnis and Norvin Richards, volume 49.
- Ortega Santos, Ivan. 2002. Overt subjects of infinitives in Spanish: Abstract agreement and the EPP. In *Selected Proceedings of the 12th Annual Graduate and Professional Symposium on Hispanic and Luso-Brazilian Literature, Language and Culture*. University of Arizona.
- Otero, Carlos Peregrín 1986. Arbitrary subjects in finite clauses. In *Generative Studies in Spanish Syntax*, ed. I. Bordelois, H. Contreras, and K. Zagona, 81–109. Dordrecht: Foris.
- Otero, Carlos Peregrín. 1999. Pronombres reflexivos y recíprocos. In *Gramática de la lengua española*, ed. Ignacio Bosque and Violeta Demonte, chapter 23, 1428–1517. Madrid: Espasa-Calpe.
- Perlmutter, D. 1970. *Deep and Surface Structure Constraints in Syntax*. New York: Holt.
- Perlmutter, D. 1978. Impersonal passives and the unaccusative hypothesis. In *Proceedings of the Fourth Annual Meeting of the Berkely Linguistic Society*, ed. J. et al Jaeger.
- Pesetsky, David. 1982. *Paths and Categories*. Doctoral Dissertation, MIT.
- Pesetsky, David. 1995. *Zero Syntax: Experiencer and Cascades*. Cambridge MA: MIT Press.
- Pesetsky, David, and Esther Torrego. 2004. Tense, Case, and the nature of syntactic categories. In *The Syntax of Time*, ed. J. Gueron and J. Lecame. MIT Press.
- Pesetsky, David, and Esther Torrego. 2007. The syntax of valuation and the interpretability of features. In *Phrasal and Clausal Architecture*, ed. S. Karimi, V. Samiiian, and W. Wilkins. Amsterdam: John Benjamins.
- Picallo, M. Carme. 1993. El contingut lèxic del pronom nul. In *Sintaxi. Teoria i Perspectives*, ed. A. Viana, 215–229. Lleida: Pagés Editors.
- Picallo, M. Carme. 2008. Gender and number in Romance. *Lingue E Linguaggio* 7:47–66.

- Piera, Carlos. 1987. Sobre la estructura de las cláusulas de infinitivo. In *Sintaxis de las Lenguas Románicas*, ed. Violeta Demonte and Marina Fernandez Lagunilla. Madrid: El Arquero.
- Pires, Acrisio. 2001. *A minimalist approach to clausal gerunds*. Ms.
- Pustejovsky, James. 1991. The syntax of event structure. In *Lexical and Conceptual Structure*, ed. B. Levin and S. Pinker, 47–81. Oxford: Blackwell.
- Raposo, Eduardo. 1985. Some asymmetries in the binding theory in Romance. *The Linguistic Review* 5:75–110.
- Raposo, Eduardo. 1987. Case theory and INF-to-COMP: The inflected infinitive in European Portuguese. *Linguistic Inquiry* 18:85–109.
- Raposo, Eduardo, and Juan Uriagereka. 1996. Indefinite se. *Natural Language and Linguistic Theory* 14:749–810.
- Reichenbach, H. 1947. *Elements of Symbolic Logic*. New York: The MacMillan Company.
- Reinhart, Tanya. 1997. Quantifier scope: How labor is divided between QR and choice functions. *Linguistics and Philosophy* 20:335–397.
- Reinhart, Tanya. 1999. Binding theory. In *The MIT Encyclopedia of the Cognitive Sciences*, ed. Robert A. Wilson and Frank C. Keil, 86–88. Cambridge Mass: MIT Press.
- Reinhart, Tanya. 2000. Strategies of anaphora resolution. In *Interface Strategies*, ed. Hans Bennis, Martin Everaert, and Eric Reuland, 295–325. Amsterdam: Royal Academy of Art and Sciences.
- Reinhart, Tanya. 2002. The Theta System: An overview. *Theoretical Linguistics* 28:229–290.
- Reinhart, Tanya. 2006. *Interface strategies: Optimal and costly computations*. MIT Press.
- Reinhart, Tanya, and Eric Reuland. 1991. Anaphors and logophors: An argument structure perspective. In *Long-distance Anaphora*, ed. Jan Koster and Eric Reuland, 283–321. Cambridge: Cambridge University Press.
- Reinhart, Tanya, and Eric Reuland. 1993. Reflexivity. *Linguistic Inquiry* 24:657–720.
- Reinhart, Tanya, and Eric Reuland. 1995. Pronouns, anaphors and Case. In *Studies in Comparative Germanic Syntax*, ed. Hubert Haider, Susan Olsen, and Sten Vikner, 241–268. Kluwer.

- Reinhart, Tanya, and Tal Siloni. 2004. Against the unaccusative analysis of reflexives. In *The Unaccusativity Puzzle: Studies on the Syntax-Lexicon Interface*, ed. Artemis Alexiadou, Elena Anagnostopoulou, and Martin Everaert. Oxford: Oxford University Press.
- Reinhart, Tanya, and Tal Siloni. 2005. The lexicon-syntax parameter: Reflexivization and other arity operations. *Linguistic Inquiry* 36:389–436.
- Reuland, Eric. 1983. Governing *-ing*. *Linguistic Inquiry* 14:101–136.
- Reuland, Eric. 2001. Primitives of binding. *Linguistic Inquiry* 32:439–492.
- Reuland, Eric. 2006a. Agreeing to bind. In *Festschrift for Henk van Riemsdijk*, ed. Hans Broekhuis and Norbert Corver.
- Reuland, Eric. 2006b. Icelandic logophoric anaphora. In *The Blackwell Companion to Syntax*, ed. Martin Everaert and Henk van Riemsdijk, volume II, chapter 33. Blackwell.
- Reuland, Eric. 2006c. Logophoricity. In *The Blackwell Companion to Syntax*, ed. Martin Everaert and Henk van Riemsdijk, volume III, chapter 38. Blackwell.
- Reuland, Eric. 2008. Anaphoric dependencies: How are they encoded? In *Reflexives and Reciprocals. Theoretical and Typological Explorations*, ed. Ekkehard König and Volker Gast. Berlin: Mouton de Gruyter.
- Reuland, Eric. 2009. Universals and variation: Encoding anaphoric dependencies. Talk given In *NORMS Workshop - Relating to Reflexives*. University of Iceland.
- Reuland, Eric, and Jan Koster. 1991. Long-distance anaphora: An overview. In *Long-distance Anaphora*, ed. Jan Koster and Eric Reuland, 1–25. Cambridge MA, Cambridge University Press.
- Rigau, Gemma. 1993. La legitimació de les construccions temporals d'infinitiu. In *Sintaxi. Teoria i Perspectives*, ed. A. Viana, 231–252. Lleida: Pagés Editors.
- Rigau, Gemma. 1994. Les propietats dels verbs pronominals. *Els Marges* 50:29–39.
- Rigau, Gemma. 1995. The properties of the temporal infinitive constructions in Catalan and Spanish. *Probus* 7:279–301.
- Rivero, Maria Luisa. 2001. On impersonal reflexives in Romance and Slavic and semantic variation. In *Romance Syntax, Semantics and L2 Acquisition*, ed. J. Camps and C.R. Wiltshire, 169–195. Amsterdam: Benjamins.
- Rizzi, Luigi. 1982. *Issues in Italian Syntax*. Dordrecht: Foris.
- Rizzi, Luigi. 1986. Null objects in Italian and the Theory of *pro*. *Linguistic Inquiry* 17:501–557.

- Rodríguez Ramalle, María Teresa. 2007. Notas sobre la reflexividad y reciprocidad en español. *Hesperia. Anuario de Filología Hispánica* X:190–206.
- Romero, Cristina. 2009. *Telicidad y Composición Aspectual*. Master's thesis, Universidad Complutense de Madrid e Instituto Universitario de Investigación Ortega y Gasset.
- Rosenbaum, Peter. 1967. *The Grammar of English Predicate Complement Constructions*. Cambridge MA: MIT Press.
- Safir, Ken. 2004. The syntax of (in)dependence. *Linguistic Inquiry Monograph* 42.
- Sánchez López, Cristina. 2002. *Las Construcciones con Se*. Madrid: Visor.
- Saussure, Ferdinand de. 1916. *Cours de Linguistique Générale*.
- Serrano Pardo, Silvia. 2008. *Un Caso Poco Estudiado de la Subordinación en Español: Sustantivas en Posición de Objeto Precedidas Por "el"*. Master's thesis, Universidad Autónoma de Madrid / Instituto Universitario de Investigación Ortega y Gasset, Madrid.
- Serrano Pardo, Silvia. forthcoming. *Estructura Informativa y Sintaxis de las Completivas Precedidas de Artículo en el Español*. Doctoral Dissertation, Universidad Autónoma de Madrid.
- Sevdali, Cristina. 2005. *Tense and Ancient Greek Infinitives*. Ms. MIT.
- Sigurdsson, Halldor A. 1991. Icelandic case-marked PRO and the licensing of lexical arguments. *Natural Language and Linguistic Theory* 9:327–363.
- Siloni, Tal. 2002. Active lexicon. *Theoretical Linguistics* 28.
- Sportiche, D. 1992. *Clitic constructions*. Ms. UCLA.
- Sportiche, D. 1998. *Partitions and Atoms of Clause Structure: Subjects, Agreement, Case and Clitics*. Routledge.
- Strozer, J. 1976. *Clitics in Spanish*. Doctoral Dissertation, UCLA.
- Tenny, Carol. 2004. *Short Distance Pronouns in Representational Noun Phrases and a Grammar of Sentience*. Ms.
- Teomiro, Ismael. 2005. *Control and Syntactic Chains: The Timing of PRO*. Master's thesis, University Utrecht.
- Teomiro, Ismael. 2008. *Control, Reflexivización e Impersonalidad: Una Aproximación Basada en las Características Léxico-Sintácticas de las Anáforas y su Procesamiento en las Interficies Lingüísticas*. Master's thesis, IUI Ortega y Gasset / Universidad Autónoma de Madrid.

- Torrego, Esther. 1984. On inversion in Spanish and some of its effects. *Linguistic Inquiry* 15:103–129.
- Torrego, Esther. 1995. From argumental to non-argumental pronouns: Spanish doubled reflexives. *Probus* 7:221–241.
- Torrego, Esther. 1998a. *The Dependencies of Objects*. Cambridge MA: MIT Press.
- Torrego, Esther. 1998b. Nominative subjects and pro-drop INFL. *Syntax* 1:206–219.
- Uriagereka, Juan. 1992. *Determiner Clitic Placement*. Ms. University of Maryland.
- Zermelo, Ernst. 1904. *Proof that every set can be well ordered*.
- Zubizarreta, Maria Luisa. 1994. El orden de palabras en español y el caso nominativo. *Nueva Revista de Filología Hispánica* VI:21–49.

Abstract

This dissertation deals with pronouns in general, and anaphors in particular. It provides a theoretical model that allows for a unified treatment of all the syntactic phenomena involving the presence of an anaphor: obligatory control and non-obligatory control structures, inherent and non-inherent reflexive verbs theme unaccusative and subject experiencer verbs that require the presence of what seems to be an anaphor in Spanish and Dutch, and arbitrary constructions such as arbitrary control, as well as Romance impersonal and passive *se/si*.

The claim is that all these syntactic phenomena have in common the presence of a SE-anaphor (henceforth SE). Except in the case of non-obligatorily controlled PRO, SE is usually deficient in referentiality due to its lack of grammatical number. It is also deficient in other φ -features such as gender, as well as in phonological realization (either a null element or a clitic). Finally, it allows for local binding iff the θ -grid of the verb has somehow been altered. The arbitrary interpretation of the SE is reached by means of existential closure of a choice function introduced by the anaphor itself.

Whereas the morphological realization of SE in control structures is cross-linguistically PRO, there are some differences in its morphological-phonological realization with reflexive, unaccusative and experiencer verbs. More specifically, whereas SE is materialized as a clitic with theme unaccusative verbs in Romance, it is not spelled out in Germanic. Furthermore, SE is materialized with subject experiencer verbs in Romance and Dutch but not in English (although SE is likewise present, as is the case with PRO). As for inherent reflexive verbs, both Spanish and Dutch spell out SE, unlike English, where SE is never spelled-out.

The present work defends the view that such cross-linguistic variation - Romance (Spanish) and Germanic (English and Dutch), in our case) - can be accounted for by resorting to processes that take place at the lexicon-syntax and syntax-phonology interfaces: *arity* operations and spell-out rules, respectively. Hence, we pursue the issue of a universal syntactic component (Chomsky 2001, 2005, 2006; Reinhart 2002; among many others). Impersonal and passive *se/si* constructions are impossible in the Germanic languages studied here because neither English nor Dutch allow SE to appear in a subject position due to its incompatibility with nominative Case, as well as the inexistence of an inherent dative SE in the lexicon.

The dissertation is structured as follows: in chapter 2 I will present the empirical data to be accounted for. I will also introduce the main guidelines of my proposal. Finally, I will sketch the theoretical framework and the model of the language upon which my analysis will be built.

In chapter 3 I will address the issue of control phenomena, as well as the nature, syntactic properties and the referential interpretation of the pronominal anaphor PRO. I will present a novel analysis of control and PRO integrated in the Minimalist Program (Chomsky, 1995 and subsequent work). After reviewing the previous literature on control in Generative Grammar, I will introduce the theoretical background upon which the analysis is built. Subsequently, I will develop the analysis, whose aim is to derive the interpretation of PRO from the Binding Theory (Chomsky 1981) as conceived by Reuland (2001, 2006, 2008), and its distribution from the interaction of Case and Phase theories (Chomsky 2001, 2005). The ultimate goal is the unified treatment of PRO and *pro* in terms of (null) SE-anaphors. The main pieces of evidence will be English *ing*-clauses and Spanish infinitives. Finally, I will present some notes on several issues that might be interesting for future lines of research.

In chapter 4 I will introduce *se/si* clitics in Romance and their counterparts in Slavic. After reviewing the most relevant literature on their status and function, I will present the core aspects of the analysis of these clitics that will be developed in subsequent chapters, including their relation with similar particles in Germanic.

Chapter 5 is devoted to showing that Spanish makes use of both complex (SELF-) and simple (SE-) anaphors. There I will demonstrate that the clitic *se* in Spanish is the morphological realization of a null SE-anaphor, and conclude that both Romance and Germanic languages form reflexives in a similar way.

SE-anaphors do not obey either Condition A or Condition B as they are formulated in Chomsky (1981). Therefore, they can be either locally bound, as I will argue is the case with inherent reflexive verbs, or non-locally bound (though this possibility is not attested in Spanish, i.e. binding in Reuland & Koster's (1991) domains 2 and 3). SELF-anaphors in Spanish (which are formed following the pattern *x+mismo*) are necessary to license the reflexive reading of non-inherent reflexive verbs. The clitics that appear with inherent reflexive verbs are SE-anaphors inserted along the syntactic derivation as a last resort mechanism in order for the derivation to converge at the C-I interface. These clitics are needed to adjust the valence (arity) of the verb and the formal requirements of syntax. In conclusion, I will show that Reinhart & Reuland's

(1993) A and B Conditions hold for English, Dutch and Spanish. Cross-linguistic variation in the occurrence of SE-anaphors with inherent reflexive verbs will be explained in terms of conditions on the spell-out of the ϕ -features of SE-anaphors (i.e. by resorting to mechanisms at the S-M interface).

In chapter 6 I will address the issue of the presence of the SE-anaphor *se* in Spanish and *zich* in Dutch with verbs other than those seen in chapter 5 (Inherent Reflexive Verbs). After dealing with some empirical and theoretical preliminary issues, I will show that there are verbs other than inherent reflexive verbs that require the insertion of SE-anaphors both in Spanish and Dutch. I will account for such a requirement, and explain why the scenarios where SE-anaphors are needed are different in Spanish and Dutch. I will show that these verbs require the presence of a SE-anaphor in English too. I will argue that The difference between the three languages lies in phonological realization: the anaphor is pronounced in Spanish and Dutch, but not in English. I will also address the question why other monadic verbs do not require or allow the presence of the pronominal particle. Finally, I will deal with alternating pronominal verbs, aspectual datives like those studied by Horn (2008), and those verbs that are ambiguous between an inchoative and a reflexive reading.

In chapter 7 I will introduce a tentative analysis of the clitic *se/si* in Romance in impersonal and passive constructions, as well as of arbitrary control across languages. The hypothesis is that, in all these constructions, there is a SE-anaphor in subject position: either PRO (in arbitrary control) or PRO' (in impersonal and passive *se/si*). I will also argue that *zich* in subject position in matrix clauses is not possible in Dutch, since this language does not materialize PRO' when it agrees with Ts (it has nominative).

The conclusions of my study will be presented in chapter 8.

Resumen

Esta tesis doctoral estudia los pronombres en general, y las anáforas en particular. También propone un modelo teórico que da cuenta de todas las construcciones sintácticas donde aparecen las anáforas en distintas lenguas. Estas construcciones son: estructuras de control (tanto de control obligatorio como de control no-obligatorio), verbos reflexivos (tanto inherentemente reflexivos como no-inherentemente reflexivos), otros verbos como los verbos inacusativos con argumento tema y los verbos de sujeto experimentante (que tanto en español como en neerlandés exigen la presencia de una partícula que resultará ser una anáfora), y finalmente, construcciones arbitrarias, como las estructuras de control arbitrario, así como las construcciones impersonales y pasivas con *se/si* en las lenguas romances.

Vamos a defender a lo largo de este trabajo que lo que estas construcciones sintácticas tienen en común es la presencia de una anáfora-SE. Este tipo de anáforas generalmente es deficiente en referencialidad (aunque no en el caso de PRO controlado de manera no-obligatoria), debido a su falta de número gramatical. Las anáforas-SE son también deficientes en otros rasgos- ϕ como el género, y frecuentemente son fonológicamente defectivas (bien son clíticos o elementos sin realización fonológica). Finalmente, permiten un ligamiento local si la rejilla temática del verbo ha sufrido algún tipo de modificación. La interpretación arbitraria de las anáforas-SE se obtiene por medio de un ligamiento existencial sobre una función de elección introducida por la anáfora.

Mientras que la realización morfológica de SE (anáfora-SE) es PRO en las estructuras de control en todas las lenguas, hay sin embargo diferencias en la realización morfológica con los verbos reflexivos, inacusativos y experimentantes. Concretamente, SE se materializa como un clítico con los verbos inacusativos de tema en las lenguas romances, pero no así en las lenguas germánicas. Por otra parte, SE se materializa con los verbos de sujeto experimentante tanto en las lenguas romances como en neerlandés, pero no en inglés (aunque defenderemos que SE está presente, como ocurre con PRO). Con respecto a los verbos reflexivos, tanto el español como el neerlandés materializan SE, a diferencia del inglés, que nunca lo materializa.

La variación observada entre las lenguas romances y germánicas —en nuestro caso, el español vs. el inglés y neerlandés— puede ser explicada mediante procesos que se dan en las interficies léxico-sintaxis (operaciones de "aridad") y sintaxis-fonología (reglas de materialización). De este modo seguiremos la hipótesis de Chomsky (2001, 2005, 2006), Reinhart (2002) y muchos otros de que el componente sintáctico es universal. Las construcciones impersonales y pasivas con *se/si* son imposibles en las lenguas germánicas estudiadas ya que tanto en inglés como en neerlandés, las anáforas-SE con materialización fonológica no pueden aparecer en posiciones de sujeto debido a la imposibilidad de marcarlas con Caso nominativo o debido a la inexistencia de anáforas-SE marcadas con caso inherente dativo en el léxico.

La tesis está estructurada como se detalla a continuación. En el capítulo 2 se presentarán los datos empíricos que serán tratados por el análisis desarrollado en la tesis. A continuación, se introducirá el marco teórico y el modelo de lenguaje sobre el que el análisis se construirá.

En el capítulo 3 se tratará la cuestión de los fenómenos de control, así como la naturaleza, las propiedades sintácticas y la interpretación referencial de la anáfora pronominal PRO. Presentaremos un análisis nuevo de control y de PRO, integrado en el Programa Minimista propuesto por Chomsky (1995, 2001). Después de revisar la bibliografía relevante sobre control dentro de la Gramática Generativa, introduciremos el marco teórico sobre el que el análisis será construido. Después desarrollaremos dicho análisis, cuya finalidad es derivar la interpretación de PRO de la Teoría del Ligamiento según la concibe Reuland (2001, 2006, 2008), y su distribución de la interacción de las teorías de Caso y de Fases. Una de las motivaciones últimas es unificar los análisis de PRO y *pro* entendiendo ambos elementos como anáforas-SE sin realización fonológica. Proporcionaremos datos del inglés (cláusulas *ing*) y del español (cláusulas de infinitivo) para apoyar el análisis de PRO defendido en este capítulo. Finalmente, presentaremos algunos comentarios sobre diversos temas que pueden ser interesantes para futuras líneas de investigación.

En el capítulo 4 haremos una introducción a los clíticos *se/si* en las lenguas romances y su correspondencia en las lenguas eslavas. En primer lugar se hará una revisión de la bibliografía más relevante sobre el estatus y la función de dichos clíticos. En segundo lugar presentaremos las hipótesis sobre estos clíticos que serán defendidas a lo largo de esta tesis, así como su relación con otras partículas similares de las lenguas germánicas.

En el capítulo 5 se mostrará que el español utiliza tanto anáforas complejas (SELF) como simples (SE). Esto es crucial, ya que argüiremos que el clítico *se* en español es la realización morfológica de una anáfora-SE, y en este sentido, defenderemos que tanto las lenguas romances como las lenguas germánicas forman reflexivos de manera similar.

Las anáforas-SE no obedecen ni a la Condición A ni a la Condición B propuestas por Chomsky (1981). De esta manera, pueden ser ligadas bien localmente, como defenderemos que ocurre con los verbos inherentemente reflexivos, o bien ligadas no-localmente (aunque esta posibilidad no se da en español). Las anáforas-SELF en español (que están formadas siguiendo el patrón $x+mismo$) son necesarias para legitimar la lectura reflexiva de los verbos no-inherentemente reflexivos. Los clíticos que aparecen con los verbos inherentemente reflexivos son anáforas-SE insertadas a lo largo de la derivación sintáctica como un mecanismo de último recurso, con el fin de que la derivación converja en la interficie con el sistema C-I. Los clíticos son necesarios para ajustar la valencia ("aridad") del verbo y los requisitos formales de la sintaxis. En conclusión, demostraremos que las Condiciones A y B de Reinhart y Reuland (1991) se aplican tanto en inglés y neerlandés como en español. La variación interlingüística que se da en la aparición de las anáforas-SE con los verbos inherentemente reflexivos será explicada en términos de condiciones de materialización de rasgos- ϕ de las anáforas-SE (es decir, mediante mecanismos en la interficie con el sistema S-M).

En el capítulo 6 trataremos la cuestión de la presencia de la anáfora-SE *se* en español y *zich* en neerlandés con otros tipos de verbos diferentes de los vistos en el capítulo 5, es decir, los verbos inherentemente reflexivos. En primer lugar daremos algunas definiciones preliminares, ciertos datos básicos que serán explicados a lo largo del capítulo, así como las cuestiones teóricas e hipótesis que serán planteadas. A continuación se describirá el marco teórico sobre el que se basa el análisis que propondremos en el capítulo. Mostraremos que hay verbos (además de los verbos inherentemente reflexivos) que requieren la presencia de anáforas-SE tanto en español como en neerlandés. Explicaremos por qué esto es así y por qué los escenarios donde las anáforas-SE son necesarias difieren entre el español y el neerlandés. Finalmente, mostraremos que el inglés también requiere anáforas-SE con los mismos tipos de verbos. La diferencia radica en si la anáfora se pronuncia (como en español y neerlandés) o no (como en inglés). A continuación, explicaremos por qué otros verbos monádicos no requieren ni toleran la presencia de la partícula pronominal. Posteriormente trataremos

los verbos pronominales alternantes, es decir, aquellos verbos que toleran pero no requieren la presencia de la partícula pronominal. También trataremos los dativos aspectuales como los que Horn (2008) estudia en el inglés, y finalmente, explicaremos por qué determinados verbos son ambiguos entre una lectura incoativa y una lectura reflexiva.

En el capítulo 7 introduciremos un análisis tentativo del clítico *se/si* en las construcciones impersonales y pasivas de las lenguas romances, así como de las estructuras de control arbitrario en distintas lenguas. La hipótesis es que en todas estas construcciones hay una anáfora-SE en posición de sujeto: bien PRO (en las construcciones de control arbitrario) o PRO' (en las construcciones impersonales y pasivas con *se/si*). También argüiremos que *zich* no es posible en neerlandés en la posición de sujeto de las cláusulas principales, ya que dicha lengua no materializa PRO' cuando concuerda con el núcleo Ts (es decir, cuando está marcada con Caso nominativo).

Finalmente presentaremos las conclusiones en el capítulo 8.

Samenvatting

Voorliggend proefschrift behandelt de pronomen in het algemeen en de anaforen in het bijzonder. Het levert een theoretisch model dat een geünificeerde verklaring geeft voor alle syntactische constructies waar anaforen in voorkomen. Bij dergelijke constructies gaat het om controleconstructies (verplichte én niet-verplichte controle), reflexieve werkwoorden (inherent en niet-inherent), andere werkwoorden zoals onaccusatieve werkwoorden en *subject experiencer*-predikaten en ten slotte willekeurige constructies zoals willekeurige controleconstructies en Romaanse onpersoonlijke en passieve *se/si* constructies.

De stelling is dat al deze syntactische constructies de aanwezigheid van een SE-anafoor gemeen hebben. Dit soort anaforen heeft gewoonlijk een gebrek aan referentialiteit (alhoewel niet in het geval van niet-verplichte controle) door het ontbreken aan een grammaticaal getal. SE-anaforen hebben ook een gebrek aan andere φ -kenmerken zoals geslacht. Verder zijn ze normaal gesproken fonologisch ontoereikend (het zijn clitics of ze worden niet uitgesproken). Ten slotte kunnen ze lokaal bonden worden als de θ -net van het werkwoord op de een of andere manier is gemodificeerd. De willekeurige interpretatie van de SE-anaforen komt voor door de existentiële sluiting van een keuzefunctie (*choice function*) die toegevoegd is door de anafoor.

Terwijl de morfologische realisatie van SE (wat op SE-anafoor wijst) PRO is in de controlconstructies door de talen heen, zijn er toch verschillen in de morfologische realisatie van SE met reflexieve, onaccusatieve en *subject-experiencer* werkwoorden. Aan de ene kant wordt SE uitgesproken als een clitic met *theme*-onaccusatieve werkwoorden in Romaanse talen maar wordt het niet uitgesproken in Germaanse talen. Aan de andere kant wordt SE uitgesproken met *subject experiencer*-predikaten in Romaanse talen en in het Nederlands maar wordt het niet uitgesproken in het Engels (hoewel SE aanwezig is op de zelfde manier als PRO aanwezig is in controleconstructies). Wat de inherent reflexieve werkwoorden betreft, zo wordt SE uitgesproken in het Spaans en Nederlands maar niet in het Engels, waar de SE-anaforen nooit lijken te worden uitgesproken.

De geconstateerde variatie door de talen heen - Romaanse talen (Spaans) en Germaanse talen (Nederlands en Engels) in ons geval - kan verklaard worden door processen die plaatsvinden in de lexicon-syntaxis ("*arity operations*") en syntaxis-fonologie interfaces. Daarom hangen we het idee van een universele syntactische component (Chomsky 2001, 2005, 2005; Reinhart 2002; onder anderen) aan. De onpersoonlijke en passieve *se/si*-constructies zijn in de tot nog toe onderzochte Germaanse talen onmogelijk omdat SE-anaforen geen casus-posities kunnen bezetten. Dit vanwege het feit dat ze of niet met nominatief kunnen worden gemarkeerd of dat er geen SE-anafoor met inherente datief in het lexicon bestaat.

Deze dissertatie is als volgt opgebouwd: in hoofdstuk 2 staan de empirische gegevens die de in dit onderzoek te ontwikkelen analyse gaat te verklaren. Vervolgens stel ik hypothesen die het onderzoek in de volgende hoofdstukken zullen leiden. Ten slotte presenteer ik zowel het theoretische kader als het model van taal waar mijn analyse zich op baseert.

In hoofdstuk 3 bespreek ik de kwestie van controle fenomenen evenals de natuur, syntactische eigenschappen en referentiële interpretatie van de pronominale anafoor PRO. Er wordt gepresenteerd een nieuwe analyse van controle en PRO die geïntegreerd is in Chomsky's *Minimalist Program* (hierna: Chomsky 1995). Na een bespreking van de bestaande literatuur over controle in de generatieve taalkunde presenteer ik het theoretische kader waarop de analyse is gebaseerd. Daarna ontwikkel ik de analyse die de interpretatie van PRO verklaart met behulp van Reulands (2001) versie van The Binding Theorie, én zijn syntactische distributie met behulp van de Casus en Fase theorieën. Één van de doelen is de analyse van PRO en *pro* te unificeren door ze allebei als SE-anaforen te beschouwen. Gegevens van Engelse *ing*-zinnen en Spaanse infinitieven worden gepresenteerd om de in deze dissertatie gepresenteerde benadering van PRO te ondersteunen. Ten slotte worden enkele opmerkingen gepresenteerd over verschillende kwesties die interessant zouden kunnen zijn voor toekomstige onderzoeklijnen.

Hoofdstuk 4 behandelt *se/si* clitics in Romaanse talen en de tegenhangers ervan in Slavische talen. Ten eerste zal ik de meest relevante literatuur over hun natuur en functie bespreken. Vervolgens stel ik mijn eigen hypothesis op over dergelijke clitics en hun functies in de verschillende constructies. Ook worden gelijksoortige partikelen in Germaanse talen behandeld.

In hoofdstuk 5 wordt geargumenteed dat Spaans gebruik maakt van zowel SE- als SELF-anaforen. Dit is erg belangrijk omdat ik wil aantonen dat de clitic *se* in het Spaans de morfologische realisatie van een nul SE-anafoor is. In deze context wil ik voorts argumenteren dat wederkerende werkwoorden in de Romaanse en Germaanse talen op een vergelijkbare manier worden gevormd. In andere woorden, ik wil aantonen dat Reinharts en Reulands (1993) A en B condities ook voor het Spaans geldig zijn.

Tevens wordt het uitgelegd dat deze taal gebruik maakt van beide SE- en SELF-anaforen. Inherent wederkerende werkwoorden ondergaan een lexicale operatie waardoor het interne argument geannuleerd wordt. Toch heeft de syntaxis altijd twee argumenten nodig. Daarom worden bepaalde clitics, die SE-anaforen zijn, ingevoegd. Dit is een soort noodmechanisme om de valentie van het werkwoord en de eisen van de syntaxis op elkaar af te stemmen. Daardoor kan de syntactische derivatie naar de CI-interface stromen. Deze clitics zijn syntactische argumenten. Echter worden ze niet als semantische argumenten geïnterpreteerd omdat ze aan de dubbele ketting voorwaarde (*Double Chain Condition*) niet voldoen. Deze voorwaarde dwingt nominale elementen zowel temporale als thematische kenmerken met het werkwoord en de temporale kernen te delen. Non-inherente reflexieve werkwoorden vereisen een SELF-anafoor, die een SE-anafoor plus een beschermend SELF element bevatten. Door de aanwezigheid van het SELF element worden de twee syntactische elementen geïnterpreteerd als twee verschillende semantische elementen aan het C-I systeem. De interpretatie van beide syntactische elementen als één zelfde element geschiedt om pragmatische redenen.

In hoofdstuk 6 wordt de kwestie van de aanwezigheid van de SE-anafoor *se* in het Spaans en de SE-anafoor *zich* in het Nederlands onderzocht met andere soorten werkwoorden dan die van hoofdstuk 5.

Ten eerste worden enkele inleidende definities behandeld samen met de basisdata die te verklaren zijn en de hypotheses. Dan wordt het theoretische kader waarop de analyse zich zal baseren, geleverd. Ik zal trachten aan te tonen dat er naast de inherent wederkerende werkwoorden ook werkwoorden bestaan die het toevoegen van SE-anaforen in het Spaans en in het Nederlands vereisen. Ik trachten te verklaren waarom dat zo is en waarom de constructies waar SE-anaforen nodig zijn, verschillend zijn in het Spaans en in het Nederlands. Tenslotte zal ik proberen aan te tonen dat ook het Engels SE-anaforen vereist met dezelfde soort werkwoorden. Het verschil is of de anafoor wel (Spaans en Nederlands) of niet (Engels) wordt uitgesproken. Vervolgens zal ik verklaren waarom andere werkwoorden met één argument geen SE-anaforen vereisen. Ik

zal de kwestie van de afwisselend pronominale werkwoorden (*nieuwe term*) behandelen. Met andere woorden, werkwoorden die de aanwezigheid van de SE-anafoor toestaan maar niet vereisen. Verder zal ik spreken over aspectuele datieven (*aspectual datives*) zoals onderzocht door Horn (2008). Ten slotte zal ik proberen te verklaren waarom bepaalde werkwoorden dubbelzinnig zijn tussen een inchoatief en een reflexieve lezing.

In hoofdstuk 7 volgt een analyse van de clitic *se/si* in Romaanse talen in onpersoonlijke en passieve constructies en willekeurige constructies door de talen heen. Mijn hypothese stelt dat er een SE-anafoor in onderwerp positie is in al deze constructies: ofwel PRO (in willekeurige constructies) of PRO' (in onpersoonlijke en passieve *se/si*). Ik zou hiermee ook willen beweren dat Nederlandse *zich* in onderwerp positie onmogelijk zijn omdat in deze taal PRO' nooit wordt uitgesproken als het met Ts (m.a.w. PRO' nominatief gemarkeerd wordt) overeenkomt.

Ten slotte volgen in hoofdstuk 8 mijn conclusies.