

A QUANTITATIVE APPROACH TO VARIABLE *SE*-MARKING  
IN SPANISH INGESTIVE VERBS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree of Doctor of  
Philosophy in the Graduate School of The Ohio State University

By

Juliana De la Mora, M. Sc

Graduate Program in Spanish and Portuguese

The Ohio State University

2011

Dissertation Committee:

Professor Scott A. Schwenter, Advisor

Professor Terrell A. Morgan

Professor Laura Wagner

Copyright by  
Juliana De la Mora  
2011

## ABSTRACT

In this dissertation, I analyze the Spanish ingestive verbs *comer* ‘eat’ and *beber/tomar* ‘drink’ that variably occur with the pronoun SE and its person/number variants. Many scholars (De Miguel and Fernandez Lagunilla 2000; Nishida 1994; Sanz 2000; Zagana 1996) have claimed that SE is an aspectual marker, and its use imposes a completive interpretation; on this view SE can only occur with telic predicates. However, it is possible to find examples that show that the alternation between SE-marked and non-SE-marked constructions is not only constrained by aspectual-related factors, but also by other factors such as the degree of individuation of the object (cf. Hopper and Thompson 1980) (1) and the (counter)expectations of the speaker (2):

- (1) Ayer vi a Sergio, (#SE) estaba comiendo unos tacos pero todavía no le servían  
‘Yesterday I saw Sergio, he was about to (#SE) eat some tacos but they hadn’t been served yet’
- (2) a. Marta se comió diez tacos de lengua  
b. #Marta comió diez tacos de lengua  
‘Marta (#SE) ate ten tongue tacos’

Maldonado (2000) and Clements (2006) have argued that the use of SE in transitive constructions increases the transitivity of the event.

Previous variationist studies of Spanish variable SE marking in motion verbs (Aaron and Torres Cacoullos 2005; Torres Cacoullos and Schwenter 2008) have revealed a set of linguistic factors that contribute to the variability. Among these are clause type, subject expression, grammatical person, tense-mood-aspect and polarity. Thus, a confluence of pragmatic and aspectual factors has been claimed to be relevant to variable SE-marking in Spanish, but qualitative studies cannot explain the myriad of factors determining this variation and their interactions. Following the variationist method, I analyzed variable SE-marking in ingestive verbs in different dialects of Spanish. My analysis of 3958 tokens from spoken and written corpora revealed non-edible/drinkable objects, highly definite and specific objects, non-human subjects, affirmative polarity, grammatical number and object position favor the SE-marked forms of *comer* and *tomar*. These results also reveal that the verb *beber* disfavors SE-marking.

Analysis of the data presented in this dissertation reveals that variable SE marking in Spanish is constrained by pragmatic factors such as animacy and counter-expectations associated with subjectivity and it is also constrained by grammatical factors associated with transitivity. I claim that the so-called aspectual SE is a non-canonical marker with a pragmatic function and a grammatical function. The pragmatic function of SE is to mark non-canonical events (either the events themselves or any of the participants) and the grammatical function of SE is to mark non-canonical transitive clauses that correspond to highly transitive clauses.

Results also reveal that the meaning of SE is determined in conjunction with the lexical/semantic properties of the verb and not by overarching properties of a generalized *SE* construction as suggested in prior qualitative research. These results, along with the findings reported by Torres Cacoullos and Schwenter (2008) and Aaron and Torres Cacoullos (2005), suggest that the choice of SE-marked constructions over non-SE-marked constructions is pragmatically driven.

## ACKNOWLEDGMENTS

Este trabajo no hubiera sido posible sin la colaboración de muchas personas. En primer lugar, mi inmensa gratitud a mi asesor Scott Schwenter, por sus invaluable críticas, cuestionamientos y comentarios. Gracias Scott por tu generosidad intelectual y bibliográfica y por la confianza que siempre tuviste en este proyecto y en mi trabajo.

Fui muy afortunada por contar en mi comité, además de con Scott Schwenter, con Terrell Morgan y Laura Wagner, dos personas y profesores extraordinarios. Gracias a Terrell Morgan por compartir todas esas otras maneras de hacer lingüística y por recordarme constantemente cuánto me apasiona lo que hago. Gracias a Laura Wagner por sus ideas y comentarios siempre brillantes y por dedicar generosamente su tiempo a este proyecto.

De una y mil maneras está presente en este trabajo John Grinstead, a quien le agradezco inmensamente todo su apoyo, trabajo y formación.

La alegría y el apoyo de la gente de la 288 fue sin duda fundamental en todo este proyecto. Gracias a Ian Tippets por compartir y contagiarme su pasión por los objetos, a Sara Sinnott por leerme y escucharme en todo momento y por el apoyo incondicional durante estos años. Gracias a Meghan Armstrong que fue cómplice de este trabajo desde el inicio y que compartió conmigo eternas discusiones sobre lo bebible y lo comible por todos los alrededores de campus. Gracias a Sandro Sessarego por su tiempo y paciencia, por las largas horas de compañía y por hacer del *think tank* un gran espacio de trabajo. Gracias a Pilar Chamorro, por las interminables, atéticas e imperfectas, charlas y carcajadas. A Daniela Salcedo, la Dany que siempre estuvo dispuesta para charlar por largo tiempo sobre cualquier tema y que sin duda hizo de la 288 un lugar más colorido.

Gracias a Mariana Vega por el apoyo, el ánimo y los cafés que me mantuvieron despierta para poder terminar esta tesis. Gracias a Evelyn Silva por todo el apoyo en todo este tiempo. Gracias a Mariana Lacunza, mi otra hermana, que me acompañó en las horas fáciles y difíciles, gracias por las eternas horas de compañía y apoyo en la cueva. Gracias a Laura Falduti por todo y por tanto, que desde allá siempre estuvo acá. Gracias a Pablo por el apoyo, la compañía y la cercanía, por compartir esto de tantas maneras y por estar siempre ahí.

Gracias Ma y Pa, por compartir el amor y la pasión al campo y por enseñarme a disfrutarlo. Gracias por todo el apoyo y por hacer de la distancia un lugar muy acogedor. Gracias a la Tati por tenerme siempre presente y por el apoyo y cariño en todo momento. Al Cam y Tat, gracias por todo, por estar siempre ahí. Gracias a la Pau por sus detalles y ánimos constantes y al Eu por el apoyo compartido también a través de la academia.

Gracias a todos los amigos y colegas de la 286 y la 288, especialmente a los que pacientemente colaboraron con sus ideas e intuiciones. Gracias a todos los amigos que desde lejos estuvieron siempre cerca.

## VITA

1998..... B.A. in Speech and Language Pathology, Colegio Superior de  
Neurolingüística y Psicopedagogía

2004..... M.Sc. in Linguistics, University of Alberta

## PUBLICATIONS

Kristin Hawkins, Kathie Houcheons, Gerardina Garita-Lacey, Juliana De la Mora and Jan Macian. 2010. Exploring the TLC: A teaching and learning center for students and instructors, *The Language Educator*, 38-43, February, 2010.

John Grinstead, Juliana De la Mora, Mariana Vega-Mendoza and Blanca Flores. 2009. An elicited production test of the optional infinitive stage in child Spanish, *Generative Approaches to Language Acquisition-North America (GALANA 2008)*, ed. by J. Crawford, K. Otaki and M. Takahashi, 36-45. Somerville, MA: Cascadilla Press.

John Grinstead, Juliana De la Mora, Amy Pratt and Blanca Flores. 2009. Temporal interface delay and root nonfinite verbs in Spanish-speaking children with specific language impairment: Evidence from the grammaticality choice task, *Hispanic Child Languages*, ed. by John Grinstead, 239–264, Amsterdam: Benjamins.

John Grinstead, Juliana De la Mora and Amy Pratt. 2007. Verbal tense and agreement in child Spanish-speakers with specific language impairment. *Proceedings of the Thirty-seventh Western Conference on Linguistics (WECOL 2007)*, ed. by Rebecca Colavin, Kathryn Cooke, Kathryn Davidson, Shin Fukuda and Alex Del Guidice, 56-67, San Diego CA: Department of Linguistics, University of California.

De la Mora, Juliana. 1999. Actas del primer encuentro sobre problemas de la enseñanza de español en México, *Español Actual*, 71:112-114, Madrid: Arco.

## FIELDS OF STUDY

Major Field: Spanish and Portuguese

Studies in: Variationist Linguistics and Pragmatics

## CONVENTIONS

A:	agent
O:	object
DO:	direct object
DOM:	differential object marking
FEM:	feminine
MASC:	masculine
NP:	noun phrase
S:	subject
SG:	singular
PL:	plural
V:	verb
VP:	verb phrase

### Corpora

CDE:	Corpus del español (Davies Corpus)
CREA:	Corpus de Referencia del Español Actual
CORDE:	Corpus Diacrónico del Español

## TABLE OF CONTENTS

ABSTRACT .....	ii
ACKNOWLEDGMENTS .....	v
VITA.....	vii
CONVENTIONS.....	ix
LIST OF TABLES.....	xiii
CHAPTER 1. INTRODUCTION.....	1
1.1 Phenomenon and goals .....	1
1.1.1 Spanish variable SE: Aspectual or transitive marker?.....	5
1.1.2 Problematic examples .....	9
1.2 Quantitative analysis of variable SE.....	14
1.2.1 Research questions and hypotheses .....	15
1.3 Transitivity and variation. A theoretical and methodological approach .....	18
1.4 Dissertation overview .....	19
CHAPTER 2. VARIABLE SE IN SPANISH .....	20
2.1 Variable SE and verbs of ingestion .....	20
2.2 Aspectual SE.....	22
2.3 Transitivity marker .....	28
2.3.1 Hopper and Thompson’s Transitivity model.....	29
2.3.2 SE as a transitivity increaser.....	34
2.3.3 Maldonado’s proposal .....	38
2.4 Transitivity marking in Spanish. Data from ingestive verbs .....	42
2.4.1 The linguistics of eating and drinking .....	43

2.4.2 Transitivity and SE marking in Spanish ingestion verbs.....	50
2.5 Variationist studies on Spanish SE.....	55
2.6 Problems with previous accounts .....	57
2.7 Conclusions .....	65
CHAPTER 3. METHODS.....	66
3.1 Variationist method .....	67
3.2 Corpus.....	68
3.3 Defining the envelope of variation .....	70
3.3.1 Exclusion criteria .....	71
3.4. Internal factors/independent variables and coding .....	75
3.4.1 Non-linguistic factors .....	77
3.4.2 Verb and clause associated factors .....	79
3.4.3 Object associated factors .....	83
3.4.4 Subject associated factors .....	93
3.5 Statistical analysis.....	95
3.6 Combining quantitative and qualitative data .....	96
3.7 Conclusions .....	97
CHAPTER 4. RESULTS.....	100
4.1 Overall statistical analysis .....	100
4.2 Statistical analysis for COMER(SE) .....	110
4.3 Statistical analysis for TOMAR(SE) .....	116
4.4 Statistical analysis for BEBER(SE).....	121
4.5 Statistical analysis of BEBER(SE) across dialects.....	125
4.5.1 BEBER(SE) in Latin American Spanish .....	126
4.5.2 BEBER(SE) in Peninsular Spanish .....	130
4.6 Qualitative analysis. Progressive and counter-expectation contexts .....	132
4.7 Conclusions .....	137
CHAPTER 5. DISCUSSION AND CONCLUSIONS.....	141
5.1 Analysis of SE as a transitivity marker: An alternative for grammar .....	144
5.2 Non-canonical events and transitivity: An alternative for pragmatics .....	148

5.3 Counter-expectations .....	153
5.4 Variable SE in Spanish .....	155
5.5 Hopper and Thompson's Transitivity model revised .....	159
5.6 Conclusions .....	160
5.7 Future research .....	163
REFERENCES .....	167

## LIST OF TABLES

Table 1. Hypotheses and predictions for SE marking .....	17
Table 2. Transitivity parameters (from Hopper and Thompson 1980).....	31
Table 3. Individuation (from Hopper and Thompson 1980) .....	33
Table 4. Transitivity parameters associated with <i>comer</i> and <i>comerse</i> (Clements 2006).....	37
Table 5. List of definite and indefinite NPs (Abbott 2004).....	85
Table 6. Linguistic factors associated with Transitivity (Hopper and Thompson 1980) ..	98
Table 7. Overall frequency of SE and non-SE-marked forms in Spanish ingestive verbs .....	101
Table 8. Internal factors contributing to the choice of SE in Spanish for all verbs.....	103
Table 9. Summary of results for the overall analysis .....	109
Table 10. Internal factors contributing to the choice of SE in Spanish for <i>comer</i> .....	111
Table 11. Summary of results for <i>comer(se)</i> .....	115
Table 12. Internal factors contributing to the choice of SE in Spanish for <i>tomar</i> .....	118
Table 13. Summary of results for <i>tomar(se)</i> .....	121
Table 14. Distribution of <i>beber(se)</i> in Latin American and Peninsular Spanish.....	122
Table 15. Internal factors contributing to the choice of SE in Spanish for <i>beber</i> .....	123
Table 16. Summary of results for <i>beber(se)</i> .....	125
Table 17. Internal factors contributing to the choice of SE in Spanish for <i>beber</i> in Latin American Spanish.....	127
Table 18. Internal factors contributing to the choice of SE in Spanish for <i>beber</i> in Peninsular Spanish.....	131
Table 19. Factor groups selected as significant for all verbs, <i>comer(se)</i> , <i>tomar(se)</i> and <i>beber(se)</i> .....	139
Table 20. Factor groups selected as significant for <i>beber(se)</i> .....	140
Table 21. Summary of results. Linguistic and non-linguistic factors that favor and disfavor SE marking .....	143

Table 22. Analysis of results based on the Transitivity Hypothesis (TH).....	147
Table 23. Contexts for <i>salir(se)</i> and <i>comer(se)</i> , <i>tomar(se)</i> and <i>beber(se)</i> .....	150
Table 24. Hypotheses, predictions and results for variable SE marking in Spanish ingestive verbs .....	158

## **CHAPTER 1. INTRODUCTION**

### **1.1 Phenomenon and goals**

In many languages reflexive pronouns can show grammatical functions other than reflexivity. This is the case in Spanish, in which the reflexive pronoun SE can occur in a wide variety of constructions with different grammatical functions. Questions regarding the semantic and syntactic status of SE have been at issue for a long time and are still an issue at present because of its wide range of functions and meanings. There have been numerous attempts to provide a unified analysis of all the manifestations of SE; however the constant discussion in the Hispanic linguistics literature about the uses of SE in Spanish shows that there is still a lot of debate around this topic. Besides its functions and meanings, there is a lot of debate about whether there is just one SE with different grammatical functions or if there are many SE's, each one with its own properties.

Most of the current literature has tried to narrow down one unique solution for the myriad theoretical problems that this pronoun presents for linguistic theories. Although there is a wide variety of analyses of SE based on different theoretical perspectives, there

is a general tendency to consider every single use of SE as a derived function from the reflexive that can be related to co-reference, impersonal constructions, inchoativity and accidentality.

In spite of the vast literature that discusses the nature, meaning and function of the SE pronoun in Spanish there is some general agreement about its grammatical characterization that includes reciprocal, reflexive, passive, middle and impersonal SE as illustrated in the following examples. The reflexive and reciprocal uses of SE are exemplified in (1) and (2) respectively; (3) is a typical case of middle voice, in which the subject or the agent corresponds with the notional object or affected object; the passive and impersonal uses of SE are illustrated in (4) and (5), respectively:

- |     |   |            |
|-----|---|------------|
| (1) | Ana <u>se</u> vio en el espejo<br>'Ana looked at herself at the mirror'                               | REFLEXIVE  |
| (2) | Los niños <u>se</u> ayudan (unos a otros)<br>'The children help each other'                           | RECIPROCAL |
| (3) | Juan <u>se</u> arrepentirá de haber tomado la decisión'<br>'Juan will regret having made the decision | MIDDLE     |
| (4) | <u>Se</u> cerraron las ventanas<br>'The windows were closed'  | PASSIVE    |
| (5) | <u>Se</u> va a las fiestas para divertirse<br>'One goes to parties to have fun'                       | IMPERSONAL |

Despite the different labels and categorizations, there is agreement on accepting the uses of SE and to a lesser extent on the idea that all these functions can be related to a reflexive pronoun with varying functions.

However there is another use of SE that is different from the previous uses, and which poses theoretical problems for the idea of one unique SE. This particular use of the pronoun has been referred to in the literature as *aspectual SE* (De Miguel and Fernández de Lagunilla 2000; Nishida 1994; Sanz and Laka 2002, Zagona 1996), *transitive SE* (Clements 2006), *middle SE/exploitation SE* (Maldonado, 1999), *reflexive dative of interest* (Alarcos Llorach 1968; D’Introno et al 2007) *ethical dative* (Gili Gaya 1960), to mention only a few. Although most of these approaches are radically different, the vast majority of the scholars agree that with this use of SE, which from now on I will call *aspectual SE*, as it is the most accepted/used term subject and object arguments are referentially independent, as in (6):<sup>1</sup>

- (6) Marta se comió unos tacos de lengua  
‘Marta SE ate some tongue tacos’

Differently from the reciprocal, passive, impersonal and middle, aspectual SE can occur with all of the SE pronoun person/number variants: *me* (1SG), *te* (2SG), *se* (3SG/PL), *nos* (1PL) and *os* (2PL), as in (7a), and the presence or absence of SE, (7a) and

---

<sup>1</sup> In the past, constructions such as (6) were not even acknowledged by Spanish Grammars (*Real Academia de la Lengua Española*, 1989; Gili Gaya, 1943) and were not considered a construction on their own but an extension of the reflexive.

(7b) respectively, does not affect the grammaticality of the sentence but the meaning, as opposed to (1) to (5) in which the use of SE is obligatory and its absence yields ungrammatical constructions or completely different interpretations:

- (7) a. Nosotros nos comimos unos tacos de lengua  
b. Nosotros ( $\emptyset$ ) comimos unos tacos de lengua  
'We (SE) ate some tongue tacos'

In spite of the extensive bibliography on Spanish SE, there is no current consensus about how this use of SE would be characterized, (i.e. what its semantic and pragmatic contributions are), since it variably occurs in similar constructions, or whether this SE has specific properties or is an extension of other uses of the reflexive pronoun in Spanish.

With some exceptions, previous studies have analyzed the uses of SE from an intuition-based and qualitative point of view, thus not being able to accurately account for the factors that constrain its variable occurrence and therefore ignoring the variation that occurs across and within varieties of Spanish.

The few quantitative studies on Spanish SE variation (Aaron 2004; Aaron and Torres-Cacoullós 2005; Torres-Cacoullós and Schwenter 2008) have revealed the importance and relevance of quantitative studies in the study of Spanish SE. Mainly, these studies have shown that it is the combination of many factors (i.e. aspectual and pragmatic) that motivates the use or non-use of SE, but unfortunately these kinds of

studies are still scarce and restricted to motion verbs, and these findings need to be incorporated into theoretical approaches to allow a better understanding of the phenomenon.

In this dissertation I propose a quantitative analysis of the variable use of SE with three verbs of ingestion in Spanish; *comer(SE)* ‘eat,’ and *beber(SE)* and *tomar(SE)* ‘drink.’ The purpose of this dissertation is to investigate the factors that constrain SE variation in Spanish and to discuss how current theoretical models can or cannot account for this particular use of SE. I also propose that an integrated solution combining conventional and context-sensitive characteristics of meaning is necessary to account for the phenomenon in question.

In the following section I will briefly discuss the most relevant theoretical frameworks and their limitations on explaining current uses of aspectual SE. I will also present my proposal and the hypotheses for this dissertation.

### **1.1.1 Spanish variable SE: Aspectual or transitive marker?**

The study of Spanish pronouns, specifically the study of reflexive pronouns, is one of the most recurring topics in the Hispanic Linguistics literature. It is not a coincidence; for ages the multiple functions of reflexives across Romance languages have caught the attention of scholars. Historically, there has been a tendency to try to narrow down the analysis of all Spanish SE as one pronoun with many different functions.

The two main approaches that currently guide the debate on the semantic/pragmatic characterization of *variable SE* are the *aspectual approach* (De Miguel 1999; Nishida 1994; Sánchez 2002; Sanz 2000; Zagona 1996) and the *transitivity approach* (Clements 2006; Maldonado 2000).

One of the most influential approaches in the literature is the one that considers SE an aspectual marker. Although the studies framed in this approach are quite different, they share some basic assumptions. It has been argued that aspectual SE is a morphological marker of accomplishments (Sanz 2000) or a culmination marker (Zagona 1996) by which the subject and object are equated to one another. Usually, aspectual SE is combined with telic predicates and, in terms of argument structure, this SE is considered to be an element with aspectual but no argumental value. It has also been claimed that its presence or absence does not change the grammatical properties of the construction and also has no consequences for the interpretation of any of the verb's arguments (Sánchez 2002). It has been considered an expletive pronoun too, which can occur with intransitive (8) and transitive verbs (9):

- (8) Ella se fue a playa  
'She SE went to de beach'
- (9) Daniel se comió una pera  
'Daniel SE ate a pear'

In terms of meaning, the presence of SE in these constructions has been related to the aspectual properties of the predicate. The aspectual SE co-occurs with perfective predicates that have specific aspectual properties and therefore is closely related to the meaning of the verb and its complements. Within this framework, it has been argued that the role of SE is to highlight a specific verbal *Aktionsart* as well as the way in which the subject is involved in the event. In these constructions, the presence of the pronoun yields a completive reading, related to perfective aspect (10), but not compatible with durative adverbial phrases such as (11):

- (10) Ana se comió la manzana en dos minutos  
'Ana SE ate the apple in two minutes'
- (11) #Ana se comió la manzana durante dos minutos  
'Ana SE ate the apple for two minutes'

The other approach to the analysis of SE considers it to be a *transitive marker*. Authors such as Garcia (1975), Maldonado (2000) and Clements (2006) have discussed the relationship between SE and the transitivity of the clause. Both authors agree on the fact that the uses of SE that have been traditionally called aspectual can be framed into a pattern in which SE highlights specific semantic properties of the verb type marked by the pronoun. While Maldonado (2000) does not necessarily claim that *se* is a transitivity marker he argues that besides highlighting the core semantic properties of the verb, SE increases the transitivity of the clause. He rejects the status of SE as an aspectual marker

and proposes that the clitic is a morpheme that “focuses on the pivotal moment of change and thus intensifies the core meaning of the event” (Maldonado 2000:155). Therefore, the role of SE with transitive verbs is to focus on the interaction established between subject and object and with intransitive verbs its role is to focus “on the most informative part of the process in which the subject participates” (Maldonado 2000:159).

García (1975) suggests that the use of the reflexive pronoun in transitive constructions lowers the verb’s transitivity and the number of arguments.

Alternatively, Clements (2006) suggests that non-anaphoric uses of SE can be analyzed under a unified model that of transitivity, as proposed by Hopper and Thompson (1980). Hopper and Thompson argue that transitivity is a property of the clauses that can be broken down into their component parts (i.e. participants, aspect, punctuality, volitionality, individuation of objects, agency, mode etc.). Hopper and Thompson (1980) propose that these component parameters of Transitivity, or parameters of Transitivity, are scalar. From here on, I will use the term Transitivity with a capital T to designate the “composite, scalar understanding of this notion” (Thompson and Hopper 2001: 28) as it was introduced in the 1980 paper. Thus on the basis of these parameters, clauses can be ranked as more or less transitive. So for example for the parameter of participants, a two or more participant clause will correspond to a high transitivity value, while a 1 participant clause will correspond with a low Transitivity value. Clements (2006) claims that the presence of SE co-varies with relatively higher Transitivity and its absence co-varies with relatively lower Transitivity. Although the author contradicts his own

argument when he claims that non-anaphoric SE can have two functions, he argues that it can have the function of reducing the transitivity of the clause (passive and middle) or it can have the function of increasing the transitivity of the clause (aspectual marker).

### 1.1.2 Problematic examples

As I mentioned above, within the aspectual approach, studies discuss only the grammatical role of SE in these constructions by arguing that it is an aspectual marker and its use can only impose a completive interpretation, and therefore it can occur only with telic predicates (De Miguel and Fernandez 2000; Nishida 1994; Sanz 2000; Zagona 1996; Bogard 2006; Bruhn de Garavito Heap and Lamarche 2002). However, data from different dialects of Spanish show that many uses of SE with verbs of ingestion such as *eat*, *drink* and *smoke* impose different interpretations: they do not necessarily encode completed events (12) and (13), and they are not restricted to telic predicates i.e. they can co-occur with durative adverbs as in (14) – (15):

- (12) Ceci, tus hijos están hermosos, me los como  
'Ceci, your kids are beautiful, I could SE eat them up'

(www.facebook.com)

- (13) Juan se comió la manzana pero dejó la mitad  
'Juan SE ate the apple, but he left half of it'<sup>2</sup>

(D'Introno et. al. 2007: 19)

Examples (14) and (15) are problematic for the aspectual approach and for the argument of the co-occurrence of SE with telic predicates. Telic predicates are compatible with time span adverbials like 'in five minutes' but are not compatible with durative adverbials like 'for five minutes' and the contrary happens for atelic predicates, which are compatible with durative adverbials such as 'for five minutes' (Dowty 1979). Based on this approach *aspectual SE* can only occur in telic predicates, therefore SE would not be compatible with durative adverbials. Examples (14) and (15) show that the presence of SE is compatible with durative adverbials:<sup>3</sup>

- (14) Nadie lo obligó, él se bebió el porrón durante/por un minuto y por eso se emborrachó  
'Nobody forced him, he SE drank the porrón for one minute, that's why he got drunk'

- (15) [Context: In a *Guinness World Records* contest, it would be felicitous to say the following:]

Durante una hora el ganador del concurso se comió 100 hot-dogs

'In a period of an hour, the winner of the contest SE ate 100 hot-dogs'

---

<sup>2</sup> This example was obtained from D'Introno et al. (2007). In their study, native speakers had to evaluate whether they would use that sentence or not. In this particular case, 75% of the participants accepted it.

<sup>3</sup> Native speakers of Mexican Spanish evaluated these two sentences as part of a pilot study of this dissertation and they considered (14) and (15) grammatical.

Examples such as (16) and (17), are also problematic for the aspectual approach. Other problematic examples are illustrated in (16) and (17). In these two examples SE is marking other features that the aspectual approach does not consider such as *volition* and *agent involvement* as in (16) and the degree of *specificity and referentiality of the object* as in (17). In (16) the volition of the agent is cancelled, as a result of a coma and the use of SE is not felicitous. In the case of (17), the absence of the *tacos* (less referential object) yields an infelicitous sentence if the SE is present (17b):

- (16) a. El paciente en coma comía la sopa a través de una sonda  
 b. #El paciente en coma se comía la sopa a través de una sonda  
 ‘The patient in a coma was SE fed through a tube’
- (17) a. Ayer vi a Sergio, estaba comiendo unos tacos pero todavía no le servían  
 b. #Ayer vi a Sergio, se estaba comiendo unos tacos pero todavía no le servían  
 ‘Yesterday I saw Sergio, he (SE) was eating some tacos, but they hadn’t served him yet’

Another example that does not fit into the categorization of aspectual SE is (18), in which the pronoun seems to be marking that something is going against the order of events –the norm is to eat soup with a spoon:

- (18) Brossa siempre dice que la gente se empeña en comerse la sopa con un tenedor  
 ‘Brossa always says that people always try to eat-SE the soup with a fork’  
 (CDE)

All these examples are problematic for an analysis that contemplates only the aspectual properties of the pronoun. It is true that the presence of SE in (12)-(18) does not change the grammatical properties of the construction, but it is also true that SE in these sentences does not necessarily yield a completive reading, as in (12), (13), (14) and (18) and, in some dialects, such as Mexican Spanish, it can occur with durative predicates (14) and (15). Another interesting property of these constructions is that in cases such as (16) in which the volition of the agent is cancelled, the use of SE is not grammatical, which suggests a close relationship between the presence of SE and volitionality. The example in (17b), shows that the definiteness of the object is not sufficient to allow the presence of SE; it also requires the object to be referential and highly individuated. Finally in (18) the presence of SE seems to be marking counter-expectations as was already found for SE/∅ variation with motion verbs (Aaron and Torres Cacoullos 2005).

As I have shown above, the aspectual approach is too narrow to account for these uses of aspectual SE in Spanish. Based on what Clements (2006) and Maldonado (1999) have proposed, a transitivity approach is more suitable for the analysis of these constructions, since it considers properties of the clause like aspect and punctuality, two of the main properties of the aspectual approach, as well as subject and object properties such as volitionality and individuation of objects to mention only a couple.

Regarding methods, another problem found in both approaches is that all the studies that have discussed the role of SE either as an aspectual marker or as a transitivity marker in these constructions are based merely on qualitative analysis of data, with only

few exceptions (Aaron and Torres Cacoullos 2005; Torres Cacoullos and Schwenter 2008; D’Introno, González and Rivas 2007).

In summary, current theoretical approaches that analyze the use of SE as an aspectual marker in Spanish present many problems:

- 1) The aspectual approach is too narrow to account for all the uses of aspectual SE in Spanish. It has been assumed that the presence of the pronoun forces a completive interpretation. However, data from many Spanish dialects show that this is not always the case and interpretations other than the completive interpretation can be obtained, thus the characterization of SE as an aspectual marker is too narrow and cannot account for many other factors that correlate with the use of SE.
- 2) The transitivity approach as a unifying approach for all the non-anaphoric uses of SE as proposed by Clements (2006) does not provide clear evidence of the relationship between the presence of SE and the transitivity of the clause; it may increase it or decrease it.
- 3) Neither the aspectual nor the transitivity approaches have considered in their analyses that SE variably appears in many constructions in Spanish. Recent variationist studies (Aaron 2003; Aaron and Torres Cacoullos 2005; Torres Cacoullos and Schwenter 2008) have shown that variable SE marking in motion verbs is constrained by both grammatical (aspect) and pragmatic factors (topicality of the subject and speaker involvement) and these factors should be considered within the theoretical approaches of Spanish aspectual SE.

4) Most previous research on SE has been exclusively qualitative and has also been limited to constructed intuition-based data, which are problematic because these kinds of studies cannot tease apart the factors that impact the use or non-use of the pronoun in spoken and written discourse.

## 1.2 Quantitative analysis of variable SE

In an attempt to solve the theoretical and methodological problems associated with the aspectual SE in Spanish, I analyzed quantitatively a specific group of verbs that variably appear with aspectual SE in two argument clauses, verbs of ingestion or consumption, *comer* ‘eat’, *tomar* ‘drink’ and *beber* ‘drink’. The restriction of the analysis to one group of verbs is due to the impossibility of studying in detail every single type of verb. The selection of this subclass of verbs is not arbitrary; ingestive verbs are the most recurrent verbs used by linguists to exemplify and illustrate variable SE (aspectual SE) in Spanish. Second, verbs such as *comer* ‘eat’ and *beber* ‘drink’ variably occur with the pronoun SE and its person/number variants *comer(se)*, *comer(me)*, *tomar(se)*, *tomar(te)*, *beber(me)*, *beber(nos)* etc. Third, ingestive verbs are high frequency verbs. Therefore it will be easy to find sufficient tokens in order to do a quantitative analysis. Finally, cross-linguistic data have revealed that this verb subclass shows a variety of properties characteristic of intransitive verbs which, as a result, are often referred to as ‘pseudo-intransitive’, ‘labile’ or ‘ambitransitive’ (Næss 2009), which is also the case for Spanish;

for this reason I consider that ingestive verbs are ideal for an analysis based on the transitivity approach. Finally, the fact that SE variably occurs in Spanish with these verbs makes them an ideal site to study via quantitative analysis through variationist methodology.

In summary, the main purpose of the study is to provide an alternative theoretical and methodological approach to the study of what traditionally has been called aspectual SE. This dissertation will also explain, from a variationist perspective, the factors that condition the distribution of SE in verbs of ingestion across different dialects of Spanish. I show that an integrated solution combining both conventional and context-sensitive characteristics of meaning is necessary to account for the phenomenon in question.

### **1.2.1 Research questions and hypotheses**

Specifically, the aim of this study is to answer to the following questions: What is the semantic/pragmatic role of SE in transitive Spanish constructions? What are the factors that determine the occurrence of SE with verbs of ingestion such as ‘eat’ and ‘drink’? What are the factors that determine the occurrence of the marked form over the unmarked form and vice versa when both forms are possible? What are the factors that guide the speakers’ choice of one form over the other and to what extent do the aspectual or transitivity properties of the clause trigger the use of SE in Spanish dialects?

For this dissertation I state the hypothesis that the SE that has been traditionally called an *aspectual marker* is a *transitivity marker*. I argue that a transitivity model such

as Hopper and Thompson's (1980) can serve as the theoretical framework to support or reject this hypothesis. With respect to verbs of ingestion, the hypotheses and predictions are the following.

First, SE-marked constructions (Marta se come la sopa 'Marta SE eats the soup') will be favored in highly transitive constructions; they will be correlated with high transitivity values of the parameters proposed by Hopper and Thompson (1980) and that non-SE-marked constructions (Marta come la sopa 'Marta eats the soup') will be correlated with low transitivity values.<sup>4</sup> If SE-marked constructions are also associated with completive aspect, they will be favored in perfective mood, by the Preterit, and disfavored by Imperfective mood.

A second hypothesis to be tested is whether 'counter-expectations' play a role in SE marking in verbs of ingestion as was already shown to be the case for motion verbs (Aaron and Torres Cacoullos 2005). Based on my previous work on the analysis in variation of *comer(se)* and *beber(se)* in Spanish (De la Mora 2009), and in Maldonado's previous work (Maldonado 1990, 2000); I predict that SE will be favored in clauses in which either the subject or the object is not prototypical (in the case of verbs of ingestion a non-prototypical subject will be one that is not animate, and a non-prototypical object will be a non-edible object or an unexpected amount of food), therefore supporting the argument of SE being a counter-expectation marker.

---

<sup>4</sup> In this dissertation I use the term SE-marked constructions, to refer to any 'transitive clause' or 'two argument clause' that occurs with SE, such as *Marta se come la sopa*, and I will use the term non-SE-marked construction to refer to transitive clauses in which the SE can occur but it does not, as in *Marta come la sopa*

Finally, on the basis of previous variationist studies (Aaron and Torres Cacoullos 2005; De la Mora 2009; Torres Cacoullos and Schwenter 2008), I predict that the use of SE will be constrained by grammatical and pragmatic factors determined in conjunction with the semantics of the verb and its complements.

Table 1 presents a summary of the hypotheses and predictions stated in this dissertation.

HYPOTHESIS	ABBREVIATION	PREDICTION
Transitivity Hypothesis	TH	SE will be favored in highly transitive clauses (associated with High Transitive Parameters) and disfavored in low transitive clauses.
Aspectual Hypothesis	AH	Perfectives will favor SE and imperfectives will disfavor SE.
Counter-expectations Hypothesis	CH	SE will be favored in clauses in which the subject or the objects are not prototypical.

**Table 1. Hypotheses and predictions for SE marking**

Besides contributing to the understanding of Spanish SE, how other meanings can be found in these constructions, and what the semantic and pragmatic conditions are for this construction and its alternations, this study will also contribute to studies of transitivity. Last but not least, this work will contribute to the semantic and cross-linguistic characterization of the linguistic properties of ‘eat’ and ‘drink’ predicates.

### **1.3 Transitivity and variation. A theoretical and methodological approach**

Based on the research questions and hypotheses stated in section 1.2, I propose a quantitative variationist analysis of SE framed within the transitivity approach as proposed by Hopper and Thompson (1980). The transitivity model is suitable for the analysis of aspectual SE for the following reasons:

- It considers transitivity as a global and gradable property of the clause that can be measured more or less in relation to a prototypical transitive clause (Hopper and Thompson 1980), therefore it allows us to identify whether SE-marked constructions are associated with high transitivity, low transitivity or if there is no relationship at all.
- Aspectual approaches can fit neatly into this model (i.e. aspect and punctuality) and should be included in order to provide a broader analysis/perspective for the analysis of aspectual SE.
- The transitivity parameters encompass properties associated with the subject and the object that seem to be crucial in SE marking as shown by previous studies and as was shown in section 1.1.2.
- Since transitivity is gradable and measurable, it is possible to test it quantitatively.

Regarding the methodological problems, I propose a variationist analysis

starting from the assumption that the use of SE in these constructions is not in free variation and the factors that determine its distribution can help us answer some of the questions regarding se-constructions that still remain unresolved.

#### **1.4 Dissertation overview**

The dissertation is organized into five chapters divided as follows. In chapter 2, I review and discuss some of the current literature on aspectual, telic and transitive SE. In this chapter I also review the relevant literature on transitivity and I discuss how this model can contribute to an alternative theoretical approach to the study of Spanish SE. Chapter 3 will be devoted to the methodology in which I will describe the data analyzed. I will define the envelope of variation of this phenomenon and describe the coding system used for the data analysis. In chapter 4, I present the results of the quantitative analysis and in chapter 5 I discuss the results and their relevance to current theories of Spanish SE and transitivity. Chapter 5 also includes the conclusions, contributions of the dissertation and possibilities for future investigation.

## CHAPTER 2. VARIABLE SE IN SPANISH

### 2.1 Variable SE and verbs of ingestion

The Spanish ingestion verbs, *comer* ‘eat’, *beber* ‘drink’ and *fumar* ‘smoke’ are among the vast number of verbs that variably occur in Spanish with SE and its person/number variants *me* (1SG), *te* (2SG), *nos* (1PL), *os* (2PL):

- (19) La primera vez que comí un taco de ojo fue una sorpresa  
‘The first time that I Ø ate an eye taco it was a surprise’  
([www.flickr.com/photos/onderyea/1384400419](http://www.flickr.com/photos/onderyea/1384400419))
- (20) fui al ISSSTE, me comí un taco de pierna de cochino enchilada  
‘I went to the ISSSTE, I ME ate a spicy pork taco’  
([talesofgreen.blogspot.com/2009\\_05\\_01\\_archive.htm](http://talesofgreen.blogspot.com/2009_05_01_archive.htm))
- (21) me incorporé al grupo, bebí el vermouth, pedí una tapa  
‘...I joined the group, I Ø drank the vermouth, I ordered a tapa’  
([www.elpais.com](http://www.elpais.com))
- (22) De un largo trago me bebí el ron que había en el vaso  
‘and in one long sip I ME drank the rum that was in the glass’

(CDE)

As was discussed in the previous chapter, the uses of SE in cases such as (20) and (22) have been traditionally called *aspectual SE*. Among the studies of aspectual SE in Spanish, the predominant kind are those studies that account for the semantic and syntactic properties of SE and the constructions in which it appears.

The co-occurrence of verbs of ingestion with reflexive morphology has been documented in other languages, for example in Dyirbal, ‘eat’ and ‘drink’ are associated with reflexive morphology indicating affectedness of the subject/agent and where no literal reflexive meaning is intended. According to Næss (2009) the presence of a reflexive pronoun with ‘eat’ in Dyirbal, does not convey a literal reflexive meaning, but rather indicates that the effect of the action on the agent is central to the meaning of the sentence (Næss 2009: 34).

In the following sections I will discuss the two main approaches that currently debate the semantic and syntactic characterization of aspectual *SE* pointing out their strengths and weaknesses. I will provide evidence showing that these two models are not exclusive, but complementary. I will show that the aspectual model can explain some of the uses of SE in transitive constructions for ingestive verbs, but that it is too narrow to account for all of them. Finally, I will provide evidence showing that the transitivity model is the most accurate approach to explain these constructions in Spanish.

## 2.2 Aspectual SE

In constructions in which SE functions as an aspectual marker, there must be agreement in person and number between the pronoun and the subject of the sentence. The aspectual marker only occurs in two argument constructions (20), the predicate must be quantized and definite (a predicate with a determiner). According to the literature it cannot occur with bare plurals (24b). It has been claimed that aspectual SE must be combined with telic predicates and for this reason it cannot co-occur with durative adverbs (25):

(23) \*Sergio se come mucho

‘Sergio SE eats a lot’

(24) a. Mi mamá nos sirve la comida y ella siempre come ensaladas

b. \*Mi mamá nos sirve la comida y ella siempre SE come ensaladas

‘Mi mother serves us the food and she always (\*SE) eats salads’

(25) \*Me comí la sopa durante todo el día

‘I ME ate the soup throughout the day’

There seems to be agreement among many scholars in considering SE as a marker of completive actions (De Miguel and Fernandez-Lagunilla 2000; Nishida 1994; Sanz 2000; Bogard 2006; Bruhn de Garavito, Heap and Lamarche 2002). In these models it has been postulated that SE is favored in telic constructions. Although most authors argue for the same properties of these constructions, there are slight differences between some of

the approaches. Nishida (1994), Zagana (1996) and Sanz (2000) claim that the predicates in SE-constructions are interpreted as telic.

Nishida (1994) argues that the SE marker in transitive constructions, namely  $SE_q$ , is an overt marker of “quantitatively delimited situations”, therefore it must contain a ‘quantitatively’ delimited direct object. This  $SE_q$  expresses quantitatively delimited situations in time or in space and marks situations that include both events and states:

All transitive sentences with  $SE_q$  have some properties in common: they model situations that are quantitatively delimited. Most typically, they express situations, dynamic or stative, in which there is a homomorphic relation between the situation and the object involved in it (Nishida 1994: 442).

In terms of the semantic properties of SE, Nishida argues that what this *se* expresses is not limited to the notion of telicity. Nishida’s model focuses on the properties of the situation. About the restrictions that SE imposes on the direct objects, she argues that they must be delimited and quantized: “those compatible with  $SE_q$  refer to entities that are quantitatively delimited, whereas those not compatible with  $SE_q$  refer to entities that do not have a clear upper bound” (Nishida 1994: 431), a property that is not required for subjects under his approach: “the quantitative delimitedness is not required of the subject of sentences with  $SE_q$ ” (Nishida 1994: 434). Specifically for ingestive verbs, Nishida (1994) proposes that when combined with SE, dynamic verbs such as *comerse* ‘to eat up’ can only form telic expressions; thus they require a direct object that

has quantized reference. Finally, regardless of perfectivity, Nishida states that in order to express quantized situations, these constructions are expected to show compatibility with tense/aspect operators.

Another author that has studied aspectual SE is Zagona (1996). She claims that the pronoun in these constructions signals a ‘special’ relationship between the subject and the object, called the predication relation, in which both arguments belong to the same place in the event. She argues that *aspectual se* explicitly marks the culmination or endpoint of the event expressed by the predicate; it is compatible only with predicates whose objects “undergo a culminating transition” (Zagona 1996: 481) therefore it occurs with definite objects but it cannot occur with bare plural objects because they are not compatible with a culmination interpretation. In the case of verbs of ingestion the activity is viewed as completed only if the change of state of the subject is completed as in (26) but not if is incomplete as in (27):

- (26) El niño (se) comió la manzana  
‘The child ate the apple (up)’  
(27) \*El niño se comió la manzana en parte  
‘The child ate up the apple partway’

(Zagona 1994: 475-476)

As De Miguel and Fernandez (2000) point out, one of the problems with Zagona’s approach is that she considers culmination and endpoint as synonyms. Under this assumption it would not be possible to explain examples such as (27) that are common in

Spanish and that are only possible if one considers culmination and endpoint as two different concepts:

- (28) Casi se comió la manzana  
*Lit* ‘He/she almost SE ate the apple’

De Miguel and Fernández (2000) claim that the “perfectiveness” notion does not account for many cases of SE use and that instead of a perfective element, SE must be considered an ‘aspectual operator’ with a quantificational nature. This argument is based on the idea that the verbal predicates can be composed of many internal phases. They adopt Pustejovsky’s theory (Pustejovsky 1991) of the structure of the events. According to Pustejovsky (1991), events can be classified based on their structure: states (unique event, no phases, i.e. ‘to have’), processes (sequence of identical events, with phases, i.e. ‘to look for’) and transitions (complex events composed by a process that ends in a state, i.e. ‘to read a book’). On this basis they claim that it is possible to focus on the different phases or stages of the event and the function of SE is to highlight that the event culminates at one point and it terminates with a change of state. Regarding the properties of the objects in these constructions, they claim that the object in these constructions must be obligatorily delimited, but unlike other approaches they do not attribute this property to the nature of the direct object *per se*, but to the restrictions that the event imposes; the event that culminates requires a delimited object. The analysis proposed is very detailed and works well for verbal forms in which there is a ‘change of state’. In

conclusion, the general idea proposed by these authors is that SE marks focus on one phase of the event structure.

Regarding the properties of the objects in SE marked predicates, Bruhn de Garavito, Heap and Lamarche (2002) claim that the presence of the clitic in these constructions indicates that the direct object must be totally affected.

Bogard (2006) calls this SE a “perfective aspectual marker”. He centers his discussion on the perfective properties of marked-SE-constructions, as motivated by the presence of the clitic, and imperfective properties of non-SE-marked constructions. In terms of its semantic interpretation, Bogard suggests that the absence of the clitic favors a durative interpretation linked to imperfective aspect, but its presence yields ungrammatical sentences such as in (29) and (30). Its presence, shown in (31) and (32), favors the completed interpretation, associated with perfective aspect:

- (29) \*Mis hijos se comieron las hamburguesas toda esta semana  
'My kids SE ate the hamburgers during all this week'
- (30) \*El señor se fue a vivir a la playa cada vez que necesitó mejorar su salud  
'The gentleman, SE went to live on the beach every time he needed to improve his health'
- (31) Mis hijos se comieron las hamburguesas  
'My kids SE ate the hamburgers'

(32) El señor se fue a vivir a la playa cuando su salud no resistió más la altura de la capital

‘The gentleman SE went to live at the beach when his health could not sustain the altitude of the capital’

(Bogard 2006: 768-769)

We find that this restriction is compatible with other verbs such as ‘to read,’ as the following example shows, in which, again, the use of SE-marked constructions is not compatible with durative adverbials, as all proponents of this approach suggest:

(33) Pedro (\*se) leyó un libro durante tres horas

‘Pedro (\*SE) read a book for three hours’

The analysis of the Spanish SE constructions is not limited to semantics, but extends also to the syntax of the constructions (Sanz 2000; Kempchinsky 2003; Mendikoetxea 1999). Sanz (2000), for example, argues that these kinds of constructions are evidence in favor of the existence of a functional projection called the Event Phrase (EP). She claims that the clitic SE in these constructions expresses accomplishments and she calls SE a *telic clitic*. This clitic occupies the head of the EP, which may host the features of telicity and punctuality. Sanz and Laxa (2002), contra Nishida (1994), claim that the SE-marker is a syntactic marker of telicity and it can only occur in sentences that express realizations. In this model the clitic is considered a functional category. For Sanz (1995) SE is a non-argument element and its occurrence is restricted by the telicity of the

sentence, thus she proposes that the presence of SE is related to the semantic information of the argument structure of the predicates.

Although there seems to be ample evidence showing that SE is an aspectual marker, I argue that this is not the only interpretation that can be obtained. There is enough evidence that shows that the contribution of SE goes further than that of an aspectual marker. In section 2.6 I will discuss some uses of SE that cannot be explained through the aspectual approach.

In the following section I will discuss an alternative theoretical model, based on the transitivity approach proposed by Hopper and Thompson (1980), and I will offer reasons for why this approach is more promising for the study of SE-constructions in Spanish.

### **2.3 Transitivity marker**

Another theoretical approach to this particular use of SE is that proposed by Clements (2006) and Maldonado (2000). Although both authors agree that SE plays a crucial role in the transitivity of the clause, their approaches to this role are quite different. Because both authors make reference to the notion of transitivity, I will discuss the main points of the model of transitivity proposed by Hopper and Thompson (1980) and that served as the theoretical model for Maldonado (2000) and Clements (2006).

### **2.3.1 Hopper and Thompson's Transitivity model**

Since the publication of works such as Hopper and Thompson's *Transitivity in Grammar and Discourse* (1980), the notion of transitivity in linguistics changed drastically and transitivity started to be conceived of as a multiscalar and gradable property of the clause.

The notion of transitivity in the past was related to a verb, which requires two argument noun phrases to form a grammatical clause. Current approaches consider transitivity a notion which relates not only to a verb and its arguments but to a number of other properties which can only be identified at the level of the clause, thus transitivity is no longer a division between transitive and intransitive verbs but a gradable notion related to the whole clause; therefore, clauses can be more or less transitive based on the clause properties.

Hopper and Thompson (1980) define transitivity in terms of the ability of the clause to display morphosyntactically transitive behavior. They claim that transitive clauses may include a morpheme that overtly codes the transitivity of the clause as well as evidence of the direct object status of the transitive patient argument of the verb. For Hopper and Thompson, transitivity is understood as a global property of an entire clause, such that an activity is carried over or transferred from an agent to a patient. Transitivity is conceived as a semantic continuum that can be analyzed in terms of ten parameters, which can be isolated, and each component can be ranked on a scale on which clauses

could be ranked as more or less transitive. On the basis of these parameters, Hopper and Thompson proposed the Transitivity Hypothesis as a language Universal: “If two clauses (a) and (b) in a language differ in that (a) is higher in Transitivity according to any of the features 1A-J, then if a concomitant grammatical or semantic difference appears elsewhere in the clause, that difference will also show (a) to be higher in Transitivity” (Hopper and Thompson 1980: 255)

Originally, transitivity ranking was conceived either as high or low, thus considering clauses with ‘high’ values for all the features as the prototype of the ‘cardinal transitivity’. In their reanalysis of the model, they claim that a clause could be marked with any number of scalar values (Thompson and Hopper 2001).

Each of the ten transitivity parameters measures the degree to which the action is transferred between participants; consequently this set of parameters allows for the characterization of clauses as more or less transitive. The transitivity parameters are summarized in Table 1.

In terms of the number of *participants* involved, highly transitive clauses are associated with the presence of two or more participants. Under the assumption that transitivity is a global property of an entire clause and that it reflects an activity that is carried over or transferred from an agent to a patient, it is assumed that “no transfer can take place unless at least two participants are involved” (Hopper and Thompson 1980: 252).

	HIGH	LOW
A. PARTICIPANTS	2 or more participants (A-O)	1 participant
B. KINESIS	Action	Non-action
C. ASPECT	Telic	Atelic
D. PUNCTUALITY	Punctual	Non-punctual
E. VOLITIONALITY	Volitional	Non-volitional
F. AFFIRMATION	Affirmative	Negative
G. MODE	Realis	Irrealis
H. AGENCY	A high in potency	A low in potency
I. AFFECTEDNESS OF O	O totally affected	O not affected
J. INDIVIDUATION OF O	O highly individuated	O non individuated

**Table 2. Transitivity parameters (from Hopper and Thompson 1980)**

The *kinesis* parameter is defined in terms of actions opposed to non-actions under the assumption that “only physical actions can be effectively transferred from one participant to another” (Guerrero Medina 2005: 350). On this basis actions, as opposed to states, can be transferred from one participant to another.

The telicity of the action is encoded by the third parameter, *aspect*; telic predicates are higher in transitivity than atelic predicates: “A predicate which specifies an endpoint or conceptual boundary is said to be telic, while one which does not is atelic.” (Hopper and Thompson 1980: 285), therefore an action viewed from its end point is more effectively transferred to a patient than one that has no endpoint. Different from

perfectivity, telicity can be determined generally by the predicate while perfectivity is a property that can only be determined or emerges in discourse.

*Punctuality* is a characteristic of actions as opposed to states. It refers to the “suddenness of an action or the absence of a clear transitional phase between onset and completion” (Hopper and Thompson 1980: 286), and distinguishes between punctual actions that are carried out with no obvious transitional phase between inception and completion and those that do have such a phase.

The fifth and eighth parameters, *volitionality* and *agency*, concern the degree of planned involvement of the agent in the activity of the verb: “Participants high in Agency can effect a transfer of an action in a way that those low in Agency cannot” (Hopper and Thompson 1980: 252). Volitionality refers to the presence of an agent acting purposefully.

The *affirmation* parameter refers to the polarity of the clause, affirmative clauses are higher in transitivity and negative clauses are lower. The *mode* parameter refers to the distinction between realis and irrealis, respectively high and low in transitivity. These terms reflect the position between “indicative and such non-assertive forms as subjunctive, optative, hypothetical, imaginary, conditional etc” (Hopper and Thompson 1980: 277).

*Affectedness of object* measures the degree of effectiveness with which the action is transferred to the object (objects can be either totally or partially affected), thus this parameter refers to the amount of intensity with which the object is affected by the verb.

Finally, *individuation of object* measures the degree of the distinctness of the object and the agent. Within this parameter, several features such as animacy and referentiality are subsumed. Individuated and non/individuated objects can be distinguished on the following basis:

INDIVIDUATED	NON-INDIVIDUATED
proper	common
human, animate	inanimate
concrete	abstract
singular	plural
count	mass
referential definite	non-referential

**Table 3. Individuation (from Hopper and Thompson 1980)**

By assuming that SE is a transitivity marker, it would be expected that SE-marked clauses show properties that correspond to highly transitive clauses. In the case of verbs of ingestion *comer(SE)*, *tomar(SE)* and *beber(SE)*, it is predicted that its occurrence will be favored in the following contexts:

- A) Participants: 2 or more participants (A-O)<sup>5</sup>
- B) Kinesis: Action
- C) Aspect (Lexical): Telic
- D) Punctuality: Punctual

---

<sup>5</sup> A refers to Agent or grammatical subject and O refers to Object

- E) Volitionality: Volitional
- F) Polarity: Affirmative
- G) Mode: Realis
- H) Agency: A high in potency
- I) Affectedness of O: O totally affected
- J) Individuation of O: O highly individuated

### **2.3.2 SE as a transitivity increaser**

According to Clements (2006), the notion of transitivity, as proposed by Hopper and Thompson (1980) can provide a unified analysis of all non-anaphoric uses of SE in Spanish. Clements points out that previous formal approaches which consider SE a valence reducer or case absorber (Bogard 2005), fail to capture other phenomena such as changes in the thematic role of the subject. Clements (2006) proposes that, “the various functions of non-anaphoric *se* are all, essentially functions of Transitivity” (Clements 2006: 236). The factors making up transitivity in Clements’ approach are the same ones that Hopper and Thompson (1980) suggest with the addition of one parameter: *individuation of subject*. Following Verhaar’s proposal (1990), Clements adds this parameter arguing that with this extension it is possible to include intransitive clauses on the transitivity scale as well. He also describes the component parts of individuation of subject and object (*individuated/non-individuated*). Thus with these values it is possible to measure the degree of transitivity of any given clause (Clements 2006). Within this definition he proposes that SE has one broad function, which is to *affect the transitivity of*

*the clause* in which it appears. SE can affect the transitivity of the clause in two different ways; it can either reduce it (as in the case of middle, passive, unaccusative and antipassive) or it can increase it (as in the case of the aspectual marker). Although Clements does not specify this, I assume that when he claims that the presence of SE co-varies with relatively higher transitivity and its absence co-varies with relatively lower transitivity, he is only referring to the aspectual marker, where the transitivity is increased, otherwise the argument is contradictory.

In the case of middle and passive constructions, Clements (2006) claims that SE acts as a valence reducer, a case absorber, neutralizer or suspender, which deprives the verb from assigning case. For middle SE constructions in Spanish, Clements (2006) states that the valency is reduced from two arguments (34a) to one (34b), where the former is higher in transitivity and the latter is lower:

(34) a. Ana enamoró a la audiencia

*Lit.* ‘Ana made the audience fall in love with her’

b. Ana se enamora fácilmente

‘Ana SE falls in love easily’

(Clements 2006: 248)

With respect to passive constructions, Clements affirms that the number of participants is reduced from two (34a) to one (34b) and the subject in the passive

constructions becomes non-agentive and as a consequence non-volitional, both considered within the low transitivity features:

- (35) (a) Marcos trabaja en la fábrica  
'Marcos works in the factory'  
(b) Se trabaja mucho en la fábrica  
'One works a lot (i.e. a lot of work goes on) in the factory'

(Clements 2006: 248)

Alternatively, Clements argues that aspectual SE increases the transitivity of the clause, this SE functions as a “marker of verbal aspect and a highlighter of foregrounded material and a marker of definiteness of different types” (Clements 2006:250). Clements discusses the verbal aspect in terms of Vendler’s (1967) classification of *Aktionsart* as achievements, accomplishment and activities. He also adds the category of states. Following this criterion Clements suggests that the *Aktionsart* can capture the difference between *dormir* and *dormirse* ‘fall asleep’; *dormir* (36a) is classified as an activity, telic and non-punctual and its marked counter part *dormirse* (36b) is classified as an achievement, being telic and punctual:

- (36) a. Marta durmió en casa de su amiga  
'Marta slept at her friend’s house'  
b. Marta se durmió en casa de su amiga  
'Marta fell asleep at her friend’s house'

Since both telicity and punctuality are considered within the high transitive parameters, this example shows how the presence of SE (47b) increases the transitivity of the clause. Clements also presents an analysis for the use of SE with verbs of ingestion such as *comer* ‘eat’ and *beber* ‘drink’. In these cases, the addition of SE makes the verb not only obligatorily transitive but also ‘sensitive to the definiteness of the object’. In terms of transitivity parameters, Clements discusses the components that are involved- aspect, and affectedness and individuation of object (Table 4, adapted from Clements, 2006).

TRANSITIVITY PARAMETERS	TRANSITIVITY VALUES	
	HIGH	LOW
	<i>comerse</i>	<i>comer</i>
ASPECT	telic	atelic
AFFECTEDNESS OF O	Object totally affected	Object not totally affected
INDIVIDUATION OF O	Individuated (count, referential, definite)	Non-individuated (mass and bare plurals, non-referential, indefinite)

**Table 4. Transitivity parameters associated with *comer* and *comerse* (Clements 2006)**

In the case of *comer* and *comerse*, all the factors that are crucial for the use of the SE-marked forms coincide with the parameters ranked for high transitivity, which provides with evidence for the role of aspectual SE as a transitive marker.

Clements also considers the variation of SE and non-SE-constructions a lexical phenomenon; he claims that there would be different lexical entries for verb pairs such as *comer/comerse* ‘eat/eat-SE’, *dormir/dormirse* ‘sleep/sleep-SE’, and *ir/irse* ‘leave/leave-SE’. Finally, Clements suggests that “the general function of non-anaphoric *se* as a marker of Transitivity may be seen as evidence that Transitivity links at least the lexicon, syntax, semantics and pragmatics and probably morphology as well” (Clements 2006: 259).

### **2.3.3 Maldonado’s proposal**

Maldonado (1999) characterizes the non-reflexive uses of *se* as middle-*se*. Maldonado (1999) frames the characterization of middle marker following Kemmer (1993) and Langacker (1991). Kemmer (1993) gives a precise definition of the middle marking as: “in general a morphosyntactic strategy for expressing an alternative conceptualization of an event in which aspects of the internal structure of the event that are less important from the point of view of the speaker are not made reference to in the utterance” (Kemmer 1993: 243). Regarding the cross-linguistic distribution and characterization of the middle marker forms, Kemmer (1993) mentions that synchronic variation in middle construction marking appears to be the norm, she also argues that middle marker verbs “tend to have corresponding unmarked forms, and the latter are usually transitive verbs denoting two-participant events” (Kemmer 1993: 21). Middle voice in Spanish carries a marker that is originally or etymologically reflexive and that, in

accordance with Langacker, indicates that the arguments of the verb are not different. Alternation of the middle marked transitive form and its unmarked counterpart are frequent in verb classes such as body care, non-translational motion, change in body posture, indirect middle, reciprocal event, spontaneous event and emotion middle types. Middle marking is semantically determined, so its use can be extended to verbs for which a corresponding transitive verb exists. Maldonado (1999) clearly establishes the differences between the reflexive and middle uses of *se* in Spanish. In the case of the reflexive construction the subject and the co-referential (in)direct object can be differentiated even when they refer to the same entity. The middles, on the other hand, involve a low degree of separateness among participants and as a consequence, a low degree of event elaboration (Kemmer 1992; Maldonado 1999). Maldonado also claims that in the middle constructions there is only either undifferentiated participant present or there are two participants that are low differentiated, as opposed to the reflexive constructions in which the participants (namely agent and patient) can be differentiated. Another distinction that Maldonado (1999) and Kemmer (1992) point out has to do with the elaboration of the event. The core function of the middle marker, according to Maldonado, is to focus on the event's crucial moment of change. Maldonado also proposes that the use of *se* marks that the object has been fully exploited. Finally Maldonado (2000) claims that in middle-*se*-marked transitive constructions "the attention will be focused on the semantic core of the verb where an undifferentiated and yet complex participant with two roles is involved in the designated process" (Maldonado

2000:160). For the characterization of all transitive verbs in middle-*se* constructions, he proposes the following generalizations:

- (i) There is only one *se* middle marker whose function is to signal a locus not differentiated from the subject and to profile the semantic core of the verb.
- (ii) The meanings of a *se* middle construction are determined by the degree of subject-object conceptual proximity in an event and by the type of interaction designated by the verb.
- (iii) In all cases, middle *se* profiles the core of the interaction designated by the verb. Thus the degree of transitivity of the construction as a whole is higher than in plain transitive construction.

According to Maldonado (2000), the Spanish SE in transitive middle constructions “highlights the maximal exploitation of the object by the clausal subject” (Maldonado 2000:153). The meaning of these constructions is determined by the degree of subject-object interaction in the event and the type of the interaction designated by the verb. Contra Clements (2006), Maldonado (2000) suggests that SE does not decrease the transitivity of the clause, but only increases it. Maldonado suggests that the function of the SE marker is to intensify and/or highlight core properties of the verbs; “the core properties of the verb is highlighted in a considerable manner” (Maldonado 2000:167). At the same time, Maldonado argues that in the case of transitive verbs, the SE marker focuses on the interaction established between the object and the subject, while in intransitive clauses it focuses on “the most informative part of the process in which the

subject participates” (Maldonado 2000: 159). The same author refers in some way to the idea of unexpected event when he claims that examples such as (37) and (38), will be pragmatically odd without SE because those special acts would expressed as if they were routine habitual event (Maldonado 2000):

- (37) Se corrió el maratón de la Ciudad de México  
‘He ran the [whole] Mexico City Marathon’
- (38) Se subió/escaló el Himalaya en dos días  
‘He went up/climbed up the Himalayas in two days’

(Maldonado 2000:175)

Maldonado explains this idea of counter-expectations in the following terms: *se* comes to be used to mark the conceptualization of the event as a counter to normal expectations (Maldonado 1999: 390-394). In terms of the syntactic properties of SE in these constructions, Maldonado (2000) claims that since SE does not occupy an argument position, it does not reduce the transitivity of the event; on the contrary “the focusing function of the clitic compresses the event selecting its nuclear properties to give them special prominence” (Maldonado 2000: 180). Opposing authors that claim that SE is an aspectual marker, Maldonado argues that completion is only one of the meanings to be obtained when the middle marker is present. His main argument is that completion derives from the internal properties of the verb (as in the case of *eat* and *drink*).

Based on Maldonado's analysis, the aspectual completed reading that traditionally has been assigned to SE is not only determined by one property but a group of features of the construction (consumption telic verb, bounded object, perfective aspect).

One of the advantages of this model is that Maldonado (2000) overtly specifies that in these constructions, one of the roles of SE is to focus on the degree of participation of the subject in the event, a concept that seems compatible with what Hopper and Thompson (1980) call volitionality and agentivity, which in the case of highly transitive clauses corresponds to an agent high in potency.

#### **2.4 Transitivity marking in Spanish. Data from ingestive verbs**

As I discussed in previous sections, the objective of this dissertation is to analyze the use and distribution of aspectual marker SE with verbs of ingestion and to determine whether or not is a transitive marker. One of the hypotheses of this work is that SE is a transitive marker and as a result its use will be favored in highly transitive constructions as measured by the ten parameters suggested by Hopper and Thompson (1980).

The variable alternation and interpretation of SE-marked constructions in Spanish, constitutes evidence for the importance and the necessity of the analysis of how transitivity works in spontaneous conversation (Thompson and Hopper 2001; Newman and Rice 2006; Vázquez Rosas 2007).

In the first part of this section I will briefly discuss the linguistics of eating and

drinking with special reference to Spanish. In the second part of this section I will discuss in terms of the Transitivity parameters, how the Transitivity model proposed by Hopper and Thompson (1980) accounts for all the cases in which the variable SE marking occurs in Spanish ingestive verbs and how this theoretical model can be adapted to the variationist methodology that I will use in this research.

#### **2.4.1 The linguistics of eating and drinking**

Across cultures and languages there are many cultural practices that surround the events of eating and drinking. Some languages mark special places to eat and drink, some other languages do have different verbs to differentiate between the transitive and intransitive uses of ‘eating’ and ‘drinking’.

Cross-linguistically, ‘eat’ and ‘drink’ have been referred in linguistic literature as “ingestive verbs” as discussed by Masica: “a small set of verbs...having in common a semantic feature of taking something into the body or mind (literally or figuratively)” (Masica 1976: 46).

Under the traditional conception of transitivity, ‘eat’ and ‘drink’ belong to what traditionally constitutes a transitive verb; a verb taking two arguments, a volitional agent and an affected patient. It is almost impossible to conceive the events of eating and drinking with an involuntary agent.

There are important syntactic and semantic properties associated with these verbs. In terms of syntax, these verbs are very peculiar because of its transitivity and in terms o

meaning these verbs are important because they are a very important source of idiomatic and figurative meanings. Verbs of ingestion across languages are very productive in terms of its figurative use; these two verbs play a very important role in the formation of figurative or idiomatic expressions across languages.

One of the reasons why these verbs have not been explored across languages is because they are a source of a lot of idiomatic expressions and this particular use of language have been marginalized in linguistics

Despite verbs like eat and drink involve two different processes in the mouth they share a lot of properties. These verbs of ingestion such as eating and drinking have some common properties; “they both involve the intake of something through the mouth and into the digestive tract; they are both vital to humans; and they are both usually enjoyed” (Newman 2009:3). These verbs are typically polysemic and frequently have extended metaphorical uses.

Another property shared by these verbs is that they can be conceptualized as both transitive or as intransitive. Based on cross-linguistics evidences, numerous scholars (Newman 2009; Næss 2007; 2009) have shown that ingestive verbs are not typical transitive predicates. It is common that across languages the events of eating and drinking are represented with two different verbs depending if they occur in a transitive or intransitive clause. While in some languages the distinction between the transitive and the intransitive uses of SE is determined by the argument structure, there are some languages that have different verbs (lexical entries) to encode transitive and intransitive eat and

drink. In the case of Spanish, the same verbal form for ‘eat’ and ‘drink’ can be used in both transitive and intransitive predicates, but there are languages in which this distinction goes further and have different lexical entries for transitive and intransitive forms. This is the case of Kiribatase and Longo that have distinct forms for transitive and intransitive eat and drink, as discussed by Newman (2009). The following example shows the transitive and intransitive alternation of ‘eat’ and ‘drink’ verb forms in Kiribatase, an Australian language:

- a. *kana* ‘eat’ transitive versus *am’arake* ‘to eat’ intransitive
- b. *nima* ‘drink’ transitive versus *mooi* ‘drink’ intransitive

(Newman 2009: 4)

The fact that these two verbs can occur as either transitive or intransitive predicates across languages, suggest that, as was mentioned lines above, these verbs are not exactly typical transitive predicates (Newman 2009).

It has also been suggested that there is a relationship between these verbs and patterns of causativization. Amberber (2009) suggests that in a number of languages it is possible to observe patterns similar to causativization in verbs of ingestion such as ‘eat’ and ‘drink’.

The importance of eating and drinking both for speakers and for linguistic reasons can be also observed cross-linguistically. For examples in Tzeltal (a Mayan language spoken in Southern Mexico), there are nine specific verbs in the eating domain (e.g. *lo*

‘eat soft things’, *we* ‘eat tortilla-like things’, *k’ux* ‘eat crunchy things’ but there is also a specific verb *tun* that means ‘eat anything’ (Brown 2008: 169).

In terms of meaning these two verbs also differ. A verb like ‘eat’ is strongly perfective in nature a property not found in ‘drink’. In the cases of the latter, there is an idea of continuity and repetitiveness not present in the former (this might be the reason why *beber* takes less the perfectivity marker than *comer*, because the lexical inherent properties of *beber* are not perfective in nature as opposed to *comer*, so the presence of *se* as a perfectivity marker with *beber* will contradict the continuity and repetitiveness meaning inherent to the verb meaning).

Metaphorical extensions and polysemy (destruction, sexual intercourse, internalization, pleasurable experience) are very frequently observed in verbs of ingestion across languages as in the case of Spanish. The figurative use of eating referring to enjoyment or satisfaction is common across languages. Another possible conceptualization of eating is that of psychological torment, distress, stress etc. present in Spanish as in the case or ‘*Me como la cabeza*’ meaning to be overthinking. Internalization is also possible as in ‘*cómete tus palabras*’ meaning something like ‘do not talk anymore’ as a way to express retraction of what has been said. According to Newman (2009), “to eat one’s words”, also involves the ‘destruction’ sense present in utterances such as ‘*el mar se comió a la playa*’, ‘the sea destroyed the beach’. It is also common that these verbs have metaphorical extensions that are sexually-related and Spanish is not an exception: ‘*me como a esa chica*’ (I will SE eat that girl) which

involves a pleasant sensory experience as well as the extension of consumption as in '*la economía se comió al país* (the economy SE ate the country), in which the meaning of eat is extended to that of consumption or destruction.

The metaphorical extensions for 'drink' are not as productive as 'eat' in Spanish, and this distinction can be explained through the semantic differences between the two verbs. Unlike eating, drinking does not involve the same kind of effect on the liquid being drunk. Thus 'drink' provides a source for continuity and repetition:

liquid which is drunk does undergo change in its movement through the digestive tract, but within the mouth and throat, which I take to be active zone of 'drink', there is no such transformation of liquid. This difference, already alluded above, has ramifications for the potential of 'eat' and 'drink' to extend to other meanings (Newman 2009: 7).

One of the metaphorical extensions of 'drink' in Spanish is that of internalization or learning as a process of emotional and intellectual satisfaction. This metaphorical extension can be observed in both marked '*bebiéndose aquellos libros en las madrugadas*' (drinking-SE those books at the sunrise) and unmarked SE forms '*porque lo justo es siempre precedente: enseñanza que bebí en el código napoleónico*' (because what is fair is always precedent: a lesson that I drank in the Napoleonic code) however, as I mentioned before, *tomar* and *beber* do not have as many semantic extensions as

*comer* in Spanish and they also show different uses across dialects, differences that I will discuss in the following paragraphs.

Although the two verbs used in Spanish for ‘drink’, *beber* and *tomar*, are both ingestive verbs, they show differences in meaning and use across and within Spanish dialects. In Latin American Spanish the most common form is *tomar*, meanwhile in Peninsular Spanish it is *beber*.<sup>6</sup> According to the Real Academia de la Lengua Española (RAE), *tomar* means to ingest alcoholic beverages in Latin American Spanish (*Am*), while the equivalent definition is given for *beber* in Peninsular Spanish.

*Tomar*. 1. Intr. *Am*. Ingerir bebidas alcohólicas.

2. tr. Comer o beber. *Tomar un desayuno, el chocolate*. U. T. C. Prnl.

*Beber*. 1. Intr. Ingerir un líquido. U. T. C. Tr.

2. intr. Hacer por vicio uso frecuente de bebidas alcohólicas.

The transitive meaning of *tomar*, can be extended to ‘eat’ as well in both dialects, ‘*tomar desayuno*’ to have breakfast or ‘*tomar un refresco*’ drink a soda.

Moreno de Alba (2003) makes reference to the common use of *tomar* in Mexican Spanish, which is associated to the ingestion of alcoholic beverages:

---

<sup>6</sup> These dialectal differences are common in online language forums, as the following example shows: “my experience has been that *tomar* is more commonly used than *beber* in the sense of *to drink*, at least among people from Latin American countries. Because of this I always suspected that *beber* was used more often in Spain, although I have very little evidence to back that up! However, when speaking about types of drink in general, one would invariably use *bebida*.” (Language forums, [www.wordreference.com](http://www.wordreference.com))

En el español mexicano, en relación con bebidas alcohólicas, existe *tomar* y *tomarse*. Este último, pronominal, siempre lleva objeto directo (“*se tomó una copa*”) y no tiene necesariamente el sentido de ‘emborracharse’. Por su parte, *tomar* puede llevar o no complemento directo (“*tomó una copa*”, “*toma mucho*”). Creo que es más usual sin objeto directo y, en tal caso, equivale precisamente a ‘beber habitualmente bebidas alcohólicas’, significado cercano ya, como se ve, a ‘emborracharse’. Pero, insisto, para llegar a ese exacto sentido, el verbo debe construirse sin pronombre y sin objeto directo. En la oración “*todos mis amigos toman*”, en el español mexicano entendemos algo así como ‘*todos mis amigos acostumbran beber bebidas alcohólicas*’.

El verbo *tomar* (y no *tomarse*) puede tener en el español mexicano (y, probablemente, en el americano en general), según el contexto, tan evidentemente el sentido de ‘beber habitualmente bebidas alcohólicas’, que de él procede el adjetivo sustantivable *tomador* (‘aficionado a la bebida’) (Moreno De Alba 2003: [www.fondodeculturaeconomica.com](http://www.fondodeculturaeconomica.com))<sup>7</sup>

---

<sup>7</sup> In Mexican Spanish both *tomar* and *tomarse* exist in relation to alcoholic beverages. The latter pronominal form always has a direct object (“she/he drank a drink”) and does not necessarily have the sense of ‘getting drunk’. On the other hand, *tomar*, can appear with or without a direct object (“she/he drank one drink”, “she/he drinks a lot”). I believe it is more common without a direct object and in that case, it has the same meaning as to “habitually drink alcoholic beverages” a meaning close to ‘getting drunk’... The verb *tomar* (but not *tomarse*) can have in Mexican Spanish, and probably in Latin American Spanish as well, the sense of habitually drinking alcoholic beverages.” My own translation.

In the case of *beber*, the most common use is found in Peninsular Spanish to refer to ‘liquid ingestion’ while in Latin America *beber* is commonly associated with the ‘ingestion of alcoholic beverages’ sense.

Although figurative uses in language have been marginalized in linguistics, I have shown in previous sections that metaphorical extensions and polysemic meanings of ‘eat’ and ‘drink’ can reveal much about language organization, categorization and specifically in the case of Spanish they are also a source for exploring language variation. For these reasons and for other methodological reasons discussed in Chapter 3, I included in all my analyses the figurative uses of eating and drinking found in the corpus.

#### **2.4.2 Transitivity and SE marking in Spanish ingestion verbs**

In this section I will briefly discuss how the Transitivity model and its parameters (Hopper and Thompson 1980) can be as a methodological and theoretical framework for the study of variable SE in Spanish.

##### *a) Participants*

Assuming that transitivity implies that an activity is carried over or transferred from an agent to a patient, it is also assumed that transfer cannot take place unless at least 2 participants are involved. This parameter presents no problem for the alternation of aspectual *se* in Spanish with verbs of ingestion, since the *se*-marked form only occurs in transitive clauses.

b) *Kinesis*

Kinesis refers to the fact that actions can be transferred from one participant to another, states cannot. This parameter is compatible with the aspectual interpretation in which it is claimed that the uses of SE cannot be extended to states. Regarding this parameter we can see that all the Spanish verbs of ingestion ‘eat’ *comer(se)* and ‘drink’ *beber(se)* are compatible with this notion and suitable for analysis.

c) *Aspect*

An action viewed from its end point, a telic action, is more effectively transferred to a patient than one not entailing an endpoint. Based on the data present here and on previous analysis (De la Mora 2009) it is possible to find SE-marked constructions that can be categorized as telic and atelic. Because of the properties of the verbs, in the case of verbs of ingestion the telic interpretation is favored over the atelic interpretation. Within this parameter two factors will be considered, telicity (lexical aspect) and perfectivity (verbal aspect). Following Hopper and Thompson’s (1980) rationale I assume that if the verbal aspect is perfective and telic, the interpretation has properties allowing the clause to be more transitive but if the aspect is imperfective and the predicate is atelic, the clause will be less transitive.

d) *Punctuality*

Actions carried out without any obvious transitional phase between the beginning and end, have more effect on their patients than those that are inherently ongoing. Within this parameter the two values punctual and non-punctual measurable on the transitive

constructions that I will analyze (i.e. durative interpretation vs. punctual). It has been discussed in the literature of ingestive predicates that the verb ‘eat’ is strongly perfective in nature (involves a change of state in the food) as opposed to ‘drink’ (does not involve the same effect on the liquid being drunk) that can be conceptualized as a continuative and repetitive event. These two lexical properties of ‘eat’ and ‘drink’ have been associated with the *punctuality parameter* and thus ‘eat’ is seen as the more transitive verb, compared with ‘drink’.

*e) Volitionality and agency*

A participant will be taken to be volitional if its involvement in the event is defined by its being volitional or sentient. The effect is more evident if the agent acts purposely. Regarding this parameter, it was discussed in previous sections that volitionality and agency are closely related to the presence of SE, see examples (16)-(17), in which the use of SE yields pragmatically infelicitous sentences when the agent’s volition is cancelled. Regarding *agency*, there is no question that participants high in agency can affect the transfer of an action in a way that those with low agency cannot. This parameter can be measured through animacy by assuming that highly animated subjects tend to act purposely.

*f) Affirmation*

More data is required to establish a conclusive argument regarding this parameter. For now, following the rationale and logic of the dissertation it is possible to measure the effects of this parameter since both forms, marked and unmarked, have no polarity

restrictions. However the prediction, following Hopper and Thompson's model is that the SE-marked constructions are going to be favored in affirmative clauses as opposed to negative.

g) *Mode*

An action, which does not occur and is non-real is less effective than one whose occurrence is actually asserted as corresponding with a real event. The distinction between a 'realis' and 'irrealis' encoding of event, can be measured by indicative and subjunctive, respectively<sup>8</sup>. The prediction of this study regarding this parameter is that SE-marked forms will be favored in clauses that code 'realis' events, since this is the value related to high transitivity.

h) *Affectedness of objects and individuation of the object*

This parameter makes reference to the distinctness of the object from the agent and from the object itself. In terms of individuation of object, individuated objects will be coded as high in transitivity while non-individuated objects will be coded as low in transitivity. The term *individuation*, in the transitivity literature, is used to refer to a number of properties that allow one to perceive a participant as an independent entity in the context (Næss 2007), the most frequently cited properties are *definiteness* and *animacy*. Because of this, I will add to the parameter 'individuation of the object' both factors, animacy and definiteness, which I will measure following Croft's animacy and

---

<sup>8</sup> More research is required to establish the difference between realis and irrealis since the opposition indicative vs. subjunctive does not always parallel the realis vs. irrealis distinction in Spanish.

definiteness hierarchies (Croft 2006).

Regarding the status of the parameter *individuation of object* in the SE-construction, a previous analysis of data from spoken and written corpora have shown that highly individuated and definite objects favor the SE-marked forms of *comerse* and *beberse* (De la Mora 2009).

i) *Counter-expectations*

A feature that is not considered under the transitivity model that seems to be marked by SE in these constructions is unexpectedness. As we can observe in example (19) when the object ingested is either ‘unexpected’, ‘abnormal’ or ‘unusual’ the marked form is favored over the unmarked form.<sup>9</sup>

I have shown that the analysis of SE as an aspectual marker can explain some but not all of its occurrences. In the previous sections I have shown how the aspectual approach is complementary to the transitivity approach and its combination can accurately explain the phenomena.

As I discussed in Section 2.2, the aspectual approach ignores semantic and pragmatic factors related to the role of the agent (*agentivity* and *volitionality*) as well as the object (*individuation, specificity and counter-expectations*) that are determinant in the use of the SE-marked form over the non-SE-marked form. Another problem is that the aspectual approach does not account for those cases in which the aspectual marker occurs and the completive reading is not obtained. Alternatively the transitivity approach

---

<sup>9</sup> This also includes metaphorical uses.

discussed in this section, can account for most the problematic uses of SE by incorporating the parameters proposed by Hopper and Thompson (1980), i.e. *aspect*, *volitionality*, *agency and individuation of object*. This approach, also allows us to interpret the role of SE as an intensifier or highlighter of the core properties of the verb and because these properties are verb-dependent, interpretations will vary.

Preliminary data presented here, along with the results of previous studies, provide evidence that the SE that has been traditionally called an *aspectual marker* would be better characterized as transitivity *marker* but a detailed quantitative analysis is still required. These data also show that the transitivity approach is a more accurate theoretical framework for explaining the variation of SE-marking in Spanish ingestive verbs. The analysis of data from corpora, following a variationist methodology, will provide more evidence for the role of SE as a transitive marker in these constructions.

## **2.5 Variationist studies on Spanish SE**

In this section I discuss some of the few quantitative studies on Spanish SE as well as the methodological and theoretical implication for the present work.

Variable SE marking is not restricted to transitive verbs in Spanish, it also occurs with intransitive motion verbs such ‘go out’ *salir(SE)* ‘go up’ *subir(SE)* and ‘go down’ *bajar(SE)*. Aaron and Torres Cocoullós (2005) studied SE variation in Mexican Spanish in the variable occurrence of SE with the verb *salir* ‘go out’. They found that both

linguistic and pragmatic factors conditioned the variation. With respect to the linguistic factors, the factors that they found as significant were first person, past tense and positive polarity. Apart from the grammatical factors, Aaron and Torres Cacoullos (2005) found that the SE form tended to be favored when there was a close relationship to the speaker and when it occurs in situations that denote deviation from social norms, a feature that is not explicitly marked in the language. Hence *salirse* occurs in four different contexts: ‘exit against obstacles or rules surreptitiously’, ‘go out abruptly’, ‘leave permanently (a group, organization or institution), and finally ‘to go out to have a good time’ (Aaron and Torres Cacoullos 2005). The authors discussed how these deviations from the social norms or ‘counter-expectations’ along with its frequency in first person and in the expression of close relation to the speaker can be taken as evidence of SE being a subjectifications marker.

In a similar variationist study, Torres Cacoullos and Schwenter (2008) investigated the factors that constrain variable SE marking with intransitive motion verbs ‘go up’ *subir(SE)* and ‘go down’ *bajar(SE)* in Mexican and Peninsular Spanish. They found that the choice of the SE marked form over the non-marked one, is determined by aspectual and pragmatic characteristics (focus on the moment of change, foregrounding, topicality of the subject, and speaker involvement). Results also show that the most important factor that determines this choice is the particular construction VERB (*subirse*, *bajarse*) + Vehicle, (50% for vehicle compared to 14% other uses). Another interesting finding in this study was the sparseness of the co-occurrence of SE constructions with

adverbials of manner, which supports the argument that SE conveys rapidity and abruptness. Within the arguments presented in favor of foregrounding they found that SE was less favorable in relative clauses, and it also co-occurs with preterit (however when preterits were broken down into perfective and imperfective no significant difference was found). Regardless of topicality and involvement, results revealed that SE occurs more frequently with first and second person than with third person (31% and 16% respectively) forms. Another interesting finding reported in this study was the low prominence of non-specific objects that was attributed to the predictability of the predicate. On these basis Torres Cacoullos and Schwenter claim that what distinguishes the SE-marked form from the non-SE-marked form is the “unexpectedness or undesirability of the situation” (Torres Cacoullos and Schwenter 2005: 1467). This study is undoubtably relevant for variationist as well as for Spanish SE studies. It not only reveals the role of pragmatics in the choice of the marked form over the non-se marked form but also shows that the specificity of the construction (verb vehicle) provides evidence that grammatical structures may be “quite lexical specific”.

## **2.6 Problems with previous accounts**

One of the main problems with the aspectual approaches discussed below is that these models reduce the interpretation of the variation of SE-marked forms and non-SE-marked forms to an event that can only be differentiated by aspectual properties (De

Miguel and Fernandez 2000) such as telicity and/or perfectivity (Zagona 1996; Bogard 2006) or marks quantitatively delimited situations (Nishida 1994). My argument and the hypothesis of this dissertation is that the contribution of SE is more complex and goes further than just aspect. In the following paragraphs I will discuss some of the interpretations obtained in the SE-marked constructions for which the aspectual marker does not necessarily account.

The first problem has to do with durative interpretations. In some varieties of Spanish, specifically in Mexican Spanish, speakers allow the durative interpretation of events that include verbs of ingestion such as ‘eat’ and ‘drink’ as the following examples show, thus contradicting the idea that these constructions are restricted to the completive interpretation. In the case of (39) and (40) it is possible to get a durative interpretation of the event, which suggests that the SE in these constructions has a different role –or more complex role- than that of aspectual marker:

- (39) a. Por las mañanas me como un poco de jamón y un poco de queso  
c. Por las mañanas como un poco de jamón y un poco de queso  
‘In the morning I (SE) eat a little bit of ham and a little bit of cheese’
- (40) Nadie lo obligó, él se bebió el porrón durante/por un minuto y por eso se emborrachó  
‘Nobody forced him, he (SE) drank the proton for a minute, and he got drunk’

Even though (39a) and (39b) are both felicitous, their interpretation is completely different. In these cases the SE is not necessarily marking a completed event or a fully

exploited object, as Maldonado (1999) suggests. The SE in this case is marking the degree of the agent involvement in the event. The same is found in (40), in which SE is marking the high degree of *involvement* and *volition* of the subject in the acts of drinking and eating. Results from an informal questionnaire reveal that speakers of Mexican Spanish accepted the form in (41) and even that for some speakers the marked form was preferred over the unmarked form in ‘casual contexts’ for example in a conversation with friends or family<sup>10</sup>.

- (41) Se bebió una cerveza en un minuto  
‘He SE drank the beer in one minute’

When a verb of ingestion such as eat or drink is modified by an adverbial phrase such as *en un minuto* ‘in one minute’, the marked-SE form is possible in the Mexican dialect.

In cases such as (42) and (43) in which the volition of the subject is cancelled either by a clinical condition, such as a coma (42) or by a voluntary act, such as a hunger strike (43), the use of SE with *comer* is not felicitous (42b) and (43b). Thus examples such as (42) and (43) are strong evidence of the role of SE in marking *volition* and *agent involvement* (cf. Hopper and Thompson 1980):

---

<sup>10</sup> Ten native speakers of Mexican Spanish answered an informal questionnaire, which included pairs of transitive sentences with and without SE. They were asked to choose the form that they would prefer in a casual conversation with friends or family. After they answer the questions, I asked them the reasons that motivated their choices.

- (42) a. El paciente en coma comía la sopa a través de una sonda  
 b. #El paciente en coma se comía la sopa a través de una sonda  
*Lit. 'The patient in a coma (#SE) ate the soup through a tube'*  
 'The patient in a coma was SE fed through a tube'
- (43) a. El preso estaba en huelga de hambre, sin embargo los celadores lo obligaban a comer, él comía el pan en contra de su voluntad  
 b. #El preso estaba en huelga de hambre, sin embargo los celadores lo obligaban a comer, él se comía el pan en contra de su voluntad  
 'The prisoner was on a hunger strike, however the prison guards forced him to eat, he (#SE) ate the bread against his will'

Examples (42) and (43) also show that despite of the 'delimited object' which is claimed to be one of the restrictions of aspectual SE (Bogard 2006; Nishida 1994) the use of the marked form *comerse* is pragmatically odd, which suggest that the presence of SE is not only limited by the presence of a 'quantized' object as Nishida (1994) claims.

The role of volition marking in these constructions is also observable with other verbs, such as 'go out' in (44). The example in (44a) shows that if the volition is 'cancelled' or 'impaired', in this case again by a clinical condition, the sentence is infelicitous; thus these examples show that SE contributes to mark this distinction as opposed to a completed event as the aspectual approach would argue:

- (44) a. #Mi hermano se salió del hospital inconsciente  
 'My brother SE got out of the hospital unconscious'

b. Mi hermano salió del hospital inconsciente  
'My brother got out of the hospital unconscious'

In the characterization of SE as an aspectual marker, Sanz (2000) claims that the uses of these forms also vary in contexts in which the sentences are used as responses to a question:

(45) A- ¿Qué comió ayer Juan para el almuerzo?

'What did John eat yesterday for lunch'

B- Comió los pasteles que quedaban

'He ate the leftover cakes'

(46) ¿Qué hizo Juan esta mañana?

'What did John do this morning'

B- **Se** comió los pasteles que le quedaban

'He SE ate the leftover cakes'

(Sanz 2000: 40)

While in (45) the question makes reference to a kind of food the subject ate, (46) gives information about an event in which the subject ate up all the leftover cakes. Sanz appeals to these examples to conclude that the object in the non-SE-marked construction is not interpreted as a measurer of the event, whereas it is indeed a measurer in the SE-marked version (46). But it is also possible to find contexts in which the SE-marked form can also be used to make reference to a kind of food as in (47):

- (47) A- ¿Qué comió ayer Juan para el almuerzo?  
 ‘What did John eat yesterday for lunch’  
 B- Se comió los seis tacos que dejaste  
 ‘He SE ate the six tacos that you left’

In (47) the use of SE is felicitous, and it refers to a very specific kind of food. I claim that specificity and preferentiality are also part of the features correlated with SE in these constructions. In this case, the properties of objects are highly specific; this property is related to what Hopper and Thompson (1980) have named the *individuation of object*. Another piece of evidence for the necessity of having highly specific objects in these constructions is found in (48). In this example the use of SE (48a) is infelicitous because it requires the object to be present (more specific and referential), while in (48b), when there is no SE marker, the presence of the object is not necessary:

- (48) a. #Ayer vi a Sergio, se estaba comiendo unas hamburguesas, pero todavía no le servían  
 b. Ayer vi a Sergio, estaba comiendo unas hamburguesas, pero todavía no le servían  
 ‘Yesterday I saw Sergio, he was about to eat (SE) some hamburgers, but he hadn’t been served yet’

This notion of referentiality is related to the notion of definiteness: “a term which refers to the relationship between a linguistic expression and the world. A noun phrase is

referential if it picks out an existing entity in the world, that is, if it refers to an actual object” (Næss 2007: 112).

The close relation between the specific characteristics of the object and the uses of *se* is again illustrated in the following examples. Data from written and spoken Spanish corpora (CREA and CORDE) shows that the use of the SE-marked form with verbs of ingestion such as ‘eat’ or ‘drink’ is preferred over the unmarked form when the object is *unexpected*, or refers to something out of the ordinary as in (49) and (50), another property for which the aspectual approach does not account:

(49) me comí toda una olla de nabos que estaua para...

‘I ME ate a whole pot of turnips’

(CORDE Corpus)

(50) No pues yo como poco cambures. Cuando se dañan me como unos diez kilos para no perderlos...

‘No, I don’t eat bananas much. When they are ripe I SE eat about 10 pounds just so they don’t go bad’

(CREA Corpus)

I found examples of this preference in almost all dialects of Spanish: the marked form is preferred if the object is *unexpected*, or refers to something out of the ordinary. It is interesting that in spite of the fact that these uses of SE (unexpected, metaphoric, etc.) have been present since the fourteenth century, the pragmatic/semantic properties that motivated these early uses have yet to be researched.

In the case of verbs of ingestion, all the metaphorical uses of ‘eat’ are marked, another interesting feature that cannot be explained via the aspectual approach:

- (51) Me sacan a la pizarra y me como la erre o me como la ene  
*Lit. ‘When I am up in front of the class I SE eat my r’s and SE eat the n’s’*  
‘When they make me go up in front of the class I don’t pronounce my r’s and n’s’  
(CREA Corpus)
- (52) Ceci, tus hijos están hermosos, me los como  
*Lit ‘Ceci, your kids are beautiful, I could eat them up’*  
([www.facebook.com](http://www.facebook.com))
- (53) El mar se come un metro de playa cada año en el Atlántico  
*Lit. ‘The sea (SE) eats one meter of the beach each year’*  
([www.clarín.com](http://www.clarín.com))

In these cases (51)-(53), the use of the unmarked form *comer* instead of *comerse* yield ungrammatical sentences, in spite of having a delimited object. The example (49), reflects the necessity of marking the verb when the direct object refers to something unexpected, metaphorical or non-edible, same as in (52) and (53). In the case of (52) and (53), I suggest that the use of SE is obligatory because of the unexpected animacy of the object and in (51) because the agent is low in animacy. Following *Comrie’s generalization* we expect (52) and (53) to be marked: “the most natural kind of transitive construction is one where the A is high in animacy and definiteness, and the P [=O] is

lower in animacy and definiteness; and any deviation from this pattern leads to a more marked construction” (Comrie 1989:128).

## **2.7 Conclusions**

I have shown in this section that there are many properties of SE-constructions that cannot be explained by the aspectual approach. First of all, this approach does not consider the role of *agentivity* and *volitionality* of the subject, which appears crucial in the election of the SE-marked form over the non-SE-marked form. Secondly it does not consider the properties of the object such as *individuation of the object* (highly specific objects are preferred in SE-marked constructions) and *counter-expectations*, which have been proven to be relevant in SE marking at least for motion verbs. Finally none of the previous theoretical approaches can accurately account for all the possible interpretations that can be obtained when the SE occurs, which as I showed are not necessarily restricted to encoding telic events.

I have also shown that a transitivity approach will be the best way to address the variable use of SE in Spanish ingestive verbs, considering all data reported by previous variationist studies on Spanish SE.

### CHAPTER 3. METHODS

As I have discussed in previous chapters, SE marking in Spanish has been overwhelmingly discussed in terms of grammatical and lexical aspect. Specifically it has been related to telicity and punctuality (e.g. De Miguel 1999; De Miguel and Fernández de Lagunilla 2000; Kempchinsky 2003; Mendikoetxea 1999; Nishida 1994; Zagona 1996) and, to a lesser extent, to transitivity (Clements 2006; Maldonado 1999). As was previously pointed out by Torres Cacoullos and Schwenter (2008), studies of Spanish SE-marking have ignored, both theoretically and methodologically, the variation that occurs in many SE-marked constructions, considering in some cases the presence of SE as an ‘optional’ or ‘expletive pronoun’ and in some others giving the false impression of a unique form-function relationship, in which there is one form for one meaning and *vice versa*. By only considering grammatical aspects, these studies have ignored other factors that have been shown to be crucial in variable SE-marking (Aaron and Torres Cacoullos 2005; Torres Cacoullos and Schwenter 2008) and that can only be investigated through a quantitative study that incorporates language use and accepts language variation as a fact.

### **3.1 Variationist method**

Variationist studies consider that language variation is not random but systematic and rule-governed (Labov 1966; Weinreich et al. 1968), therefore speakers have linguistic options available and these options are not randomly chosen.

Within this approach it is assumed that language varies in a systematic way and whenever there is a choice among two or more linguistic alternatives and the choice may be influenced by several external or internal factors, the distribution of the variants can be quantitatively modeled using statistical techniques (Young and Bailey 1996; Sankoff 1988). The variationist method not only attempts to find and determine the usage patterns in the relative frequency of co-occurrence of linguistic forms but it also accounts for the variability by allowing variable rules to apply in different contexts at different probability levels. Thus the goal of variationist analyses is not to discover individual occurrences or overall rates of occurrence, but structured patterns of variability in the data investigated (Tagliamonte 2001:731).

It is from a variationist perspective (Labov 1969; Milroy and Gordon 2003; Sankoff 1988) that I will address the variation in the use of SE with verbs of ingestion. The approach adopted here presents a means to test quantitatively the assumptions made in previous literature, mainly the use of SE as an aspectual, telicity and/or transitivity marker, in order to investigate the factors that favor the occurrence of SE and its

grammatical contribution. I tackle the question of the contribution of grammar and pragmatics to speakers' choice of SE-marked forms of *comer*, *beber* and *tomar* via quantitative analysis of variation between marked (*comerse*, *beberse*, *tomarse*) and non-marked forms (*comer*, *beber*, *tomar*) in corpora of naturally occurring language.

### **3.2 Corpus**

Corpus linguistics has become an autonomous methodological paradigm within linguistics and has been a very useful resource in variationist studies. The use of corpora in linguistic research enables the analysis of naturally occurring data.

As previous studies have shown, language variation can be explored through corpus-analysis. The use of corpus-based analysis provides a means of handling large amounts of data and it also allows for keeping track of many contextual factors. At the same time it is an excellent source of naturally occurring data, which is crucial for understanding specific constraints on language variation.

Because this study is a first attempt to identify the factors that constrain the variation, it was necessary to obtain a “big picture” of the phenomena that includes a wide variety of contexts as well as a representative sample (in terms of frequency) suitable for generalizations. For these reasons I decided to include linguistic data from spoken and written language in order to capture the subtleties of the variation in a representative data sample while considering a wide variety of contexts.

The data analyzed in this dissertation was extracted from *Corpus del Español (CDE)*, created by Mark Davies (Davies 2002). This corpus is free, and is available online through <http://www.corpusdelespanol.org/>. It consists of 100 million words from more than 20,000 Spanish texts from the 1200s to the 1900s and naturally occurring spoken data (taken from interviews). The CDE includes data from different Spanish dialects, including Argentina, Bolivia, Colombia, Costa Rica, Cuba, Chile, Dominican Republic, Guatemala, Honduras, México, Paraguay, Peru, Puerto Rico, Spain, Uruguay, Venezuela and Spanish from the U.S. This corpus allows users to perform advanced searches based on part of speech, lemma, collocates, synonyms, and frequency in different time periods and genres, “the interface allows you to search for exact words or phrases, wildcards, lemmas, part of speech, or any combinations of these. You can search for surrounding words (collocates) within a ten-word window” (Davies 2002).

Another advantage of the interface used in the CDE, is that it allows you to limit searches by frequency and compare the frequency of words, phrases, and grammatical constructions by register and by historical period. Because the objective of this dissertation is to analyze present-day variation, I only included tokens from the 20<sup>th</sup> century in my searches and subsequent analysis. Another factor considered was dialectal variation. Since Spanish SE-variation is a phenomenon that has been observed in all Spanish dialects, I decided to include data from all Latin American Spanish dialects as well as Peninsular Spanish.

### 3.3 Defining the envelope of variation

If we assume that a variable reflects choices among forms that speakers make, it is crucial to define the contexts in which the choices are possible and the contexts in which they are not. Determining this variable context, also known as the envelope of variation, is essential in variationist studies because it affects all subsequent analyses and results, as well as the interpretation of the results. In order to determine the envelope of variation we need to take into account not only the form that we are analyzing but also the forms with which that form varies. This consideration, known as the principle of accountability, requires inclusion of all contexts where variation is possible and exclusion from the analysis of all contexts where variation cannot occur. This principle specifies that variationist analyses must account for every case where the variable element can occur in the relevant environments, as these are defined (Labov 1972; cf. Milroy and Gordon 2003; Poplack and Tagliamonte 2001).

In the present study, it was necessary to identify and delimit the variable contexts in which SE-marking occurs with verbs of ingestion, specifically *comer* and *tomar/beber*. The envelope of variation was determined inductively. Despite the fact that these verbs do not variably occur with SE in its intransitive forms, the CDE does not support searches by either transitive or intransitive form of the verb, thus it was necessary to extract all the occurrences, both transitive and intransitive and then to extract manually only the transitive uses of *comer(se)*, *beber(se)* and *tomar(se)* and all their person/number variants.

For each token, I extracted the necessary context for the analysis of the variable. In cases in which the context available in the corpora was not enough to determine any of the factors that I was analyzing, the tokens were excluded from analysis.

Many scholars have claimed that one of the syntactic restrictions of the use of aspectual SE in these constructions is that the predicate must be ‘quantized’ (Nishida 1996), must occur with a determiner and cannot occur with bare plurals. However, I found enough occurrences in the corpus that show that SE occurs even with bare plurals (54)-(56). Following the assumption that every variant that is part of the variable context must be taken in to account, I decided to include in the data all transitive uses of the three verbs even if the predicates were formed with bare plurals, since variation occurs:

- (54) Después se comía *bisteques enormes* con tres o cuatro huevos fritos  
‘Afterwards she/he SE ate enormous steaks along with three or four fried eggs’
- (55) Estos no se han comido *níscalos* como debían de comérsele  
‘They have not SE eaten mushrooms as they should have eaten them’
- (56) penetró en la plaza, donde antes había *pasto*, que las vacas se habían comido  
‘he/she got into the plaza, where there was *grass* before, that the cows SE had eaten’

(CDE)

### 3.3.1 Exclusion criteria

As part of the task of determining the envelope of variation, it was necessary to establish all the cases that were excluded from the analysis. After extracting all the

possible uses of the three verbs, I excluded all the intransitive uses that cannot be SE marked and therefore variation does not occur (58):

- (57) Comí bien  
‘I ate well’
- (58) \*Me comí bien  
‘I SE ate well’

As mentioned in Chapter 1, the pronoun SE has several grammatical functions besides the reflexive use, such as impersonal, passive and reciprocal. Although SE can occur with verbs of ingestion in impersonal, reflexive and reciprocal constructions, all these uses of SE were excluded from the analysis because no variation is manifested in such contexts; either the pronoun is obligatory or its absence yields a completely different interpretation. Therefore, all the impersonal, reflexive and reciprocal uses of SE with the verbs eat and drink, such as (59)-(61) were excluded for the analysis:

- (59) Sabes que aquí en la posguerra se ha comido boniato  
‘you know that here in post-war times people SE have eaten sweet potato’
- (60) la carne se lleva de otros países o no se come carne  
‘the meat is brought from other countries or meat is not SE eaten’
- (61) sentirías el mismo terror que ver a miles de insectos copulando y comiéndose los unos a los otros  
‘you will feel the same terror as if you were looking at thousands of insects copulating and SE eating each other’

(CDE)

In addition to intransitive uses, reflexive, impersonal and reciprocal SEs idiomatic expressions were not considered as potential cases of variation, since they are lexicalized with respect to marking or not, and variation does not occur. The decision of whether one expression should be excluded from the envelope of variation was based on whether that particular expression that could potentially be considered an idiom or, sometimes, a discourse marker, showed variation in the corpora used for this dissertation, the most recurrent idiomatic phrase, was *con su pan se lo coma* (good for him/her, it's his/her loss, too bad) and it was always excluded since there is no variation of SE-marking.

(62) algún lector no tiene sentido del humor, pues con su amargo pan se lo coma

‘some readers do not have a sense of humor, too bad for him/her’

(63) Diego, con su pan se lo coma; haz tu negocio y el del Rey

‘Diego, his/her loss, do your own business and the king’s’

(CDE)

At the same time, all tokens in which any of these verbs were used with a different meaning than that of ingestion were excluded from the data used for the analysis. This task of exclusion was carried out by hand since there is no tool in the corpus that allows one to restrict by both grammatical functions (verbs) and semantic content at the same time:

(64) ¿Qué efecto le produce cuando lee esto? - No lo tomo como algo triste

‘What effect does this have on you when you read this? I don’t take it as something sad’

(65) Entonces tomó otro taxi y entonces justo en la puerta del [...

‘And then he/she took another cab and just at the door of the [...’

(66) Hace veinte! Precisamente; cuando nació la beba

‘Twenty years ago. Precisely when the baby was born’

(67) se marca en la escritura con punto o punto y coma

‘it is marked in writing with a period or a semi-colon’

(68) desemboca en edema cerebral, trastornos neurológicos y coma profundo

‘it may end in brain swelling, neurological disorders and deep coma’

(CDE)

All the tokens that included *comer*, *beber* and *tomar* in the ingestion sense, either literal or metaphoric, and that occur in two-argument predicates (A-O) were included in the analysis. The following examples illustrate the uses of verbs of ingestion with meanings like to have (69), to finish something (70) and to destroy (74).

(69) Viene a buscarme y toma el desayuno conmigo

‘He/she comes to pick me up and he/she has breakfast with me’

(70) se comió una fortuna. Fue perdiendo todo, hasta las ganas de salir

‘he/she SE ate a fortune. He/she was loosing everything, even the’

(71) los...los del Caribe...se comían algunas vocales, pero nunca las consonantes

‘people in the Caribbean SE ate some vowels, but never consonants’

(72) a la patria? " dice: " He venido a servirla pero no a comérmela oiga "

‘the Country?’” He/she says:” I have come to serve the country not to SE eat it’

(CDE)

- (73) Uyy uyy uyy... yo si me como a esa mujer. Está muy, pero muy bien esta chava  
*Lit.* uy uy uy...I would SE eat that woman. She looks so so fine  
'I would chat up that woman. She looks so so fine'  
(www.gamers.com)
- (74) que vivimos en sociedad y, la sociedad, ciertamente, come mucho la moral  
'we live in society, and certainly society eats morality'  
(CDE)

### 3.4. Internal factors/independent variables and coding

After determining the envelope of variation, I extracted the tokens in which variation can occur and excluded all contexts where the use of SE is invariant, indeterminable or in which the verb refers to something different from ingestion. The tokens extracted from the remaining variable contexts were coded for factors that potentially motivate SE-marking. These factors have been derived from the literature on aspectual SE-marking and from previous studies that have reported SE-variation in motion verbs in Spanish. Because of the nature of the hypothesis, it was necessary to code for many factors in order to determine the role that aspect, telicity, perfectiveness and transitivity were playing in the variation of SE with verbs of ingestion.

Because of the quantitative nature of this endeavor, the numerous hypotheses, including the proposed features and factors relevant to SE-marking, needed to be operationalized for the analysis. In regression analyses these individual factors are the independent variables that may potentially affect the dependent variable (SE-marking).

Every token, both marked and unmarked, has to be coded for the same independent variables in order to be interpretable by the variable rule program Goldvarb X (Sankoff et al. 2005). The independent variables considered for this analysis are described in the following sections. Following the variationist method and terminology, these variables are organized into factor groups in the Goldvarb program, and the various options within each factor group are called factors. Once the tokens were extracted and filtered, and the factors were determined, data were recorded using the Excel software program for uniformity, organization and ease of use in managing the data.

The *modus operandi* of quantitative analysis of language variation is to cast as wide a net as possible regarding the potentially motivating factors of a phenomenon. In other words, it is methodologically preferable to be maximally inclusive in the possible motivating factors in order to capture the relevant factors. In the following section I detail all of the initial factor groups and their factors used for coding so that my methodology and considerations are completely transparent and explicit. In Chapter 5 I discuss my analyses and detail the circumstances and decisions that motivated recoding the data through collapsing and eliminating factor groups and factors.

Every token extracted from the corpus was coded for linguistic and non-linguistic factors. The linguistic variable was coded in a way that is consistent with the goals of my study; a) to determine whether the factors that favor the occurrence of SE-marked forms, b) to find out if these factors are related to aspectual structure, as can be inferred from previous theoretical models that suggest that SE is an aspectual marker, c) to determine

whether these factors are transitivity-associated and finally d) to investigate whether the factors that determine the variation of SE in verbs of consumption are similar or parallel to those found in motion verbs.

Although the extra-linguistic information available from the CDE is not extensive there are two factors that can provide significant information of the variation of SE, namely dialect and register. Since this information was already available from the corpus, I coded the tokens for these two extra-linguistic factors.

Linguistic factors were determined based on previous variationist studies and operationalization of Hopper and Thompson's (1980) Transitivity model. As I have discussed in previous sections not all the transitivity parameters proposed by Hopper and Thompson (1980) are transparent with respect to the ways in which one can measure them. However, I operationalized these parameters using one or more than one factor, susceptible to measurement as detailed in the following paragraphs.

### **3.4.1 Non-linguistic factors**

#### *3.4.1.1 Dialect*

Although the variation in SE-marking on verbs of ingestion is observed in all dialects of Spanish, it has not been determined whether there are dialects in which one form is more widespread than the other. A goal of this dissertation is also to examine whether the parameters that determine the variation are the same across dialects. In order

to do so, I coded all the tokens extracted from the corpus by country. The data extracted from the CDE for this analysis, comes from 17 different dialects of Spanish: Argentina, Bolivia, Colombia, Costa Rica, Cuba, Chile, Dominican Republic, Guatemala, Honduras, México, Paraguay, Peru, Puerto Rico, Spain, Uruguay, Venezuela and Spanish from the U.S. One of the main problems of the data obtained from the corpus, was that there were not enough tokens for each Latin American dialect, therefore I grouped all the data from Latin American countries in one group and data from Peninsular Spanish in another group, hence I had two factors for this factor group.

Another problem of this corpus, discovered after retrieving the tokens, is that only in the oral examples is the source (in this case the dialect) included. For all the tokens extracted from written registers, it was necessary to find the author and then determine the dialect by which that text should be classified. This search also served to corroborate that all tokens extracted corresponded to the 20th Century, which was then double-checked through a search for the year in which authors were born.

#### *3.4.1.2 Register*

Variable use of SE in consumption verbs is not restricted to a specific register- it occurs in both spoken and written Spanish. A previous study on SE variation in Spanish (Torres Cacoullós and Schwenter 2008) revealed that the factor that contributes the most in SE marking for *subirse* and *bajarse* was register (called “medium” in their paper). Based on these results and assuming that written registers could also shed some light on the behavior of SE in different registers and that the comparison between oral and written

registers will reveal important information about the variation of the pronoun across registers, I included tokens from both oral and written registers. In order to determine whether register was a factor that favors or disfavor the use of SE in verbs of ingestion, as was previously found in motion verbs, I coded all tokens for either *spoken* or *written* register, based on the information found in the corpus. Following previous results the prediction for this factor was that SE should be more likely to occur in oral register.

### **3.4.2 Verb and clause associated factors**

#### *3.4.2.1 Verb type*

The verbs *comer*, *tomar* and *beber* were chosen because they are all verbs of ingestion (therefore they are semantically similar), variation occurs with the three of them, they are high frequency verbs so it is easy to find them in the corpus. This variable was included to determine whether the use of SE in these verbs is lexically determined and constrained.

#### *3.4.2.2 Clause type*

In this dissertation transitivity is considered a global property of a clause and is conceived as a gradient feature, which can be measured as more or less, in relation to a prototypical transitive clause that describes an activity that is carried over or transferred from an agent to a patient (Hopper and Thompson 1980). At the clause level, transitivity could be affected by clause type, negation, mood and aspect. With respect to clause type, it has been argued that declarative clauses are more transitive than interrogative clauses,

since it is more likely that an action is transferred from an agent to a patient in a declarative as opposed to an interrogative in which the transfer does not necessarily occur. For this factor group I coded tokens as either a *declarative* or *interrogative* clause, predicting, under the assumptions of the Transitivity Hypothesis, that SE-marking will be favored in declarative clauses.

#### 3.4.2.3 Polarity

Another factor that influences transitivity at the clause level is negation. The factor polarity corresponds to the *Affirmation* parameter in Hopper and Thompson's model (1980). In their model, affirmative clauses are associated with high transitivity and negative clauses are associated with low transitivity, thus the effect of negation is to reduce the transitivity of the clause. Polarity has also been associated with foregrounding.

Another piece of justification for considering negative clauses as low in transitivity is given by Næss (2007). Her approach correlates the effect of negation in the transitivity of the clause with the properties of the object of a negated sentence, hence negation affects the transitivity of the clause through a modification of the semantic features that the arguments would have in the affirmative sentences. As a consequence, the semantic opposition between these arguments decreases and as a result, the transitivity of the clause decreases too. Næss argues that if there are no perceivable effects, then the subject cannot be characterized as an agent since there is no actual event that an agent could be an agent of, thus if John didn't eat the apple, then he is not perceived as 'agentive'. This analysis explains why negation may have an impact on the

transitivity of the clause:

... if a subject of a negated clause is a Frustrative rather than an Agent, then there is no longer semantic opposition between its arguments, and the clause is therefore less semantically transitive. As far as its contribution to the transitivity of the clause is concerned, then, we can conceive of negation as an operator changing the value of certain features of the arguments of the clause, with a consequent reduction in the semantic opposition between these arguments (Næss 2007:115).

Under this assumption I coded each token based on the polarity of the clause, as *affirmative* or *negative*, predicting that affirmative clauses will be more transitive than negative clauses. The occurrences that were coded for negative polarity included those negated with *no* (not), as well as other negators such as *nunca* (never), *ni* (nor) and *jamás* (never ever):

- (75) Sí que hay unas cosas que... que *no comería* nunca más  
‘There are some things that...that I would *never eat* again’
- (76) Ellos *nunca tomaban* chocolate al desayuno sino mazamorra  
‘They *never drank* chocolate for breakfast but rather mazamorra’
- (77) Yo alcohol *no he bebido* nunca  
‘I have *never drunk* alcohol’

(CDE)

By hypothesizing that SE is associated with high transitivity, it is predicted that SE will be favored in affirmative clauses.

#### 3.4.2.4 Aspect

The following factors were included in this factor group: *preterit*, *imperfect*, *commands*, *present*, *others*. Previous variationist studies (Torres Cacoullos and Schwenter 2008) have shown that preterit favors the SE-marking construction as opposed to non-preterit contexts, which disfavor it. Considering results from previous variationist studies and assuming that SE is associated with perfective aspect, I predict that SE will be favored in perfectives (i.e. preterit) and disfavored in imperfectives (i.e. imperfect).

#### 3.4.2.5 Mood

The *mode* parameter refers to the distinction between realis and irrealis mode, respectively high and low in transitivity. According to Næss, these terms reflect the position between “indicative and such non-assertive forms as subjunctive, optative, hypothetical, imaginary, conditional etc” (Hopper and Thompson 1980: 277). Although the binary distinction between realis and irrealis in Spanish is not clearly determined by mood, I operationalized this factor by coding tokens for *indicative* (78), (79), *imperative* (80) and (81), *subjunctive* (82) and *conditional* (83) as illustrated in the following examples:

- (78) Lo dijo suavemente, colocándose la pastillas en la lengua y bebiendo un sorbo.  
‘He/she said it softly, putting the pills on his/her tongue and drinking a sip’

- (79) En un snack bar comió dos hamburguesas y bebió un vaso de leche  
 ‘At a snack bar he/she ate two hamburgers and drank a glass of milk’
- (80) pero en adelante no comas otra cosa que huevos cocidos  
 ‘from now on, do not eat anything but hard-boiled eggs’
- (81) Rosa, ¿quieres un cachito de esto? ¿Eh? No, cómetelo tú  
 ‘Rosa, do you want a piece of this? What? No, you eat-SE it’
- (82) Antes, por la mañana pues era raro que ella se tomara el aperitivo, ¿verdad?  
 ‘Before, in the morning it was rare that she would SE drink an aperitif, right?’
- (83) Se bebería de un buche media botella de coñac  
 ‘he/she would SE drink half bottle of cognac in one gulp’

(CDE)

### 3.4.3 Object associated factors

Following the Transitivity model proposed by Hopper and Thompson, there are two parameters associated with objects: *affectedness of object and individuation of object*.

*Affectedness of object* measures the degree of effectiveness with which the action is transferred to the object (objects can be either totally or partially affected) thus this parameter refers to the amount of intensity with which the object is affected by the verb.

Finally, *individuation of object* measures the degree of the distinctness of the object and the agent. Within this parameter several features such as animacy and referentiality are subsumed. Individuated and non/individuated objects can be distinguished on the following basis:

INDIVIDUATED	NON-INDIVIDUATED
proper	common
human, animate	inanimate
concrete	abstract
singular	plural
count	mass
referential	non-referential
definite	

In order to operationalize these two parameters, I coded objects for *definiteness*, *specificity*, *object number*, *object position* and finally *edibility/drinkability*. Although object position and edibility are not considered in the Transitivity model, I included them to test the counter-expectation hypothesis (CH), predicting that SE will be favored in preverbal object clauses and when occurring with non-edible objects.

#### 3.4.3.1 *Definiteness and specificity*

Definiteness is another factor that plays a crucial role across languages. Object properties tend to be marked in many different ways. For example in Amwi, an Austro-Asiatic language spoken in Bangladesh, the usual word order is VSO, but when the object is indefinite it occurs immediately after the verb (Hopper and Thompson 1980: 259). In Spanish, definiteness and specificity have been also related to differential object marking (DOM) (Leonetti 2004; Tippets 2010; von Heusinger and Kaiser 2003, 2005).

In order to determine if the presence of SE was associated with highly specific and affected Objects, I coded DO as definite and indefinite according to the following criteria (Abbott 2004):

DEFINITE NPs	INDEFINITE NPs
Pronouns, demonstratives, definite descriptions, possessive NPs, proper names, NPs with a universal quantifier as a determiner (i.e. <i>cada, todos</i> ), the null determiner understood generically	Bare NPs (understood existentially, any (polarity sensitive any, free choice any), no, most, a/an ( <i>un, uno, una, unas</i> ), several, a few, many and indefinite this

**Table 5. List of definite and indefinite NPs (Abbott 2004)**

Examples (84-86) show examples of tokens with objects coded as definite and examples (87-90) show indefinites.

*Definite:*

(84) Y se los ha comido *todos*. Masticando como una moledora. Qué hambre terrible, ¿no?

‘And he has eaten *all* them up. Chewing like a machine. What a terrible hunger, no?’

(85) pero a juzgar por sus demostraciones de equilibrio ya se habían bebido *la bodega entera*.

‘judging by his balance they already SE had drunk the whole bodega’

(86) Pues me comeré *su almuerzo*

‘I will ME eat *his lunch*’

(CDE)

*Indefinite:*

- (87) También comen *calamares, anguilas, pequeños tiburones y algas*  
'They also eat *squids, eels, small sharks and seaweeds*'
- (88) Ya no bebía *una cerveza* con cada comida  
'I did not drink *a beer* with every single food'
- (89) No me importa, me como *cualquier cosa*  
'I don't care, I ME eat *whatever I found*'
- (90) preparó *unas quesadillas* con tocino que comió de pie en la cocina,  
'he made *some quesadillas* with bacon that he ate standing in the kitchen'

(CDE)

Another factor group associated with object properties is *specificity*. This factor group consisted of two factors: *specific* and *non-specific*. The categorization of objects followed Shain's proposal (2009): specific objects consist of uniquely identifiable referents from the speaker's perspective based on the discourse context. The measurement of specificity used in this investigation, depends on the de dicto vs. de re distinction. If referents are selected from and stand apart from a particular set, they are specific. Non-specific referents are those that are not unique among the members of their sets.

*Non-specific:*

- (91) no recuerdo *cuántas* nos habíamos tomado ya  
'I cannot remember *how many* NOS have we drunk yet'
- (92) *lo mejor de la olla* se lo ha comido  
'the best of the dish, she has SE eaten it'

- (93) Ya nos tomaremos una copa para celebrarlo  
 ‘We will NOS have a drink to celebrate’
- (94) habla... discutiendo en una mesa, tomándose sus copas y discutiendo de...  
 ‘he talks, arguing in the table, drinking-SE his drinks and arguing about’
- (CDE)

*Specific:*

- (95) antes de venirse a su casa se bebía dos botellitas de vino; comía,  
 ‘before he got home he SE drink two bottles of wine, he ate’
- (96) él se comía dos cambures, o un... racimo de uvas, o..  
 ‘he SE ate two plantains or a bunch of grapes’
- (CDE)

*3.4.3.2 Object number*

Many authors have discussed how singular objects can be more individuated than plural objects. Hopper and Thompson (1980) argue that highly individuated objects are higher in transitivity than non-individuated objects, and singular objects are more individuated than plural objects. Following this claim I coded objects in my corpus as *singular* and *plural*, predicting that singular objects will favor SE-marking over plurals (see Table 3, Individuation).

Cases such as *un par de tragos* ‘a couple of drinks’ were coded in the plural forms. All occurrences of objects in interrogative clauses were coded as singular objects (97) unless there was a plural NP or a plural WH-word (98) and (99):

- (97) ¿Qué comiste?

- ‘What did you eat?’
- (98) ¿*Cuántas* copas nos tomamos?
- ‘How many drinks did we NOS drink?’
- (99) ¿*a cuáles* de sus amigos nativos se habían comido?
- ‘Which one of your native friends had they SE eaten?’

(CDE)

### 3.4.3.3 Object position

Another factor group that was included in the coding was object position. Although there is no Transitivity parameter related to object position. However, since one of the hypotheses predicts that SE is associated with counter-expectations or non-prototypical features (either in discourse or grammar), I decided to include object position as a possible site to test this hypothesis. For this factor group, I assumed, as has been suggested by many authors, that topics can be syntactically and intonationally marked in languages. Because of the nature of the data analyzed, I only considered the syntactic test for topicality. In this dissertation, I assume (following von Heusinger and Kaiser 2007) that a left moved direct object is topical, thus topicality is expressed or at least detectable in texts, by left-dislocation (von Heusinger 2008). Following canonical word order in Spanish, I predict that if SE is a marker of counter-expectation or a marker of a non-prototypical form, it will be favored in clauses that have left-dislocated objects and therefore do not have the canonical word order SVO but OVS. For this factor, I coded tokens for *preverbal* or *postverbal* objects. In cases such as *quiero beberlo* (I want to drink it) in which the object was pronominalized, the object was coded as preverbal

assuming that the referent was previously mentioned in the discourse. For questions, objects were coded also by their position with respect to the verb. Wh-questions such as (100) were coded as preverbal, while questions like (101) were coded as post-verbal objects:

(100) *¿Qué quiere usted tomar?*

*‘What do you want to drink?’*

(101) *¿Quiere tomar algo?*

*‘Do you want to drink something?’*

(CDE)

Questions in which there is a pronoun and an NP were coded as preverbal (102), since it is assumed that the referent was previously activated in discourse:

(102) *¿Se lo bebe así el vinagre?*

*‘SE do you drink vinegar like this?’*

(CDE)

#### *3.4.3.4 Edibility*

Within the transitive model there is no parameter that can account for the specific nature of the object. Previous studies have shown that the type of event that the verb refers to matters for variation. For example, Torres Cacoullos and Schwenter (2008) discuss the counter-expectations of an event as a context in which SE is more likely to occur with motion verbs. Based on these results with motion verbs and in previous

studies with ingestive predicates, I decided to code for the specific nature of the object. De la Mora (2009) showed that SE is favored when it co-occurs with objects that are ‘unexpected’ which in the case of verbs of ingestion will be objects that are not edible or that people usually do not eat. Since there is no parameter associated with this property, I coded the tokens in my corpus following Leach’s (1964) characterization of edible substances. Leach (1964) argues that edible substances can be characterized as follows (1) edible substances that are recognized as food and consumed as part of the normal diet, (2) edible substances that are recognized as possible food, but that are prohibited or else allowed to be eaten only under special conditions and (3) edible substances that by culture and language are not recognized as food as all. For this factor I coded for *edible/drinkable* and *non-edible/drinkable objects*. Since I was interested in the difference between edible and non-edible objects I grouped categories (2) and (3) and I coded all objects that fit into that category as non-edible and all objects that fit into category (1) were coded as edible objects:

*Edible/drinkable objects*

- Edible substances that are recognized as food and consumed as part of the normal diet:

(103) después de haber parado en la desembocadura del Aconcagua para comerse  
*un tomate fresco*

‘after we stopped on top of the Aconcagua to eat-SE a fresh tomato’

- (104) comía *el arroz*, lo recogía en el tenedor con precisión  
 ‘he ate *the rice*, he grabbed it on his fork with precision’
- (105) *fruta y también carne* procuro no comer carne  
 ‘*fruit and also meat*, I try not to eat meat’
- (106) pues me dijo, levantándose: - bébetete pronto esa leche. La bebí de un sorbo  
 ‘so, he said, drink-TE that milk soon. I drank it in a gulp’
- (107) Me paro en un sitio, me tomo un café, y estoy un tiempo solo  
 ‘I stop at a place, I ME-drink a coffee and I spend some time alone’
- (108) cuando todo había concluido y bebíamos nuestra copa en el otro extremo de la ciudad  
 ‘when everything had finished and we drank our glass on the other side of the city’
- (109) Nos bebemos tres o cuatro copas seguidas y – y NO veas lo que alegra eso  
 ‘Nos we drink three or four drinks in a row and- and you don’t know how happy that makes us’

(CDE)

*Unexpected objects (non-edible food or drinks)*

- Edible substances that are recognized as possible food, but that are prohibited or allowed to be eaten only under special conditions.

- (110) Como los niños de los comedores escolares que tomaban *leche adulterada*  
 ‘Like the kids in the school cafeterias that drank *adulterated milk*’
- (111) Pero ahora no le quitaba la vista al vaso. Cuando menos había bebido una *docena de cubas*  
 ‘He kept staring at the glass. At least he has drunk a dozen of rum and cokes’

▪ Edible substances that by cultural practices and language are not recognized as food as all:

(112) de los cocodrilos, que se comen a los hombres enteros.

‘of the crocodiles, that SE eat entire men’

(113) los indios tarahumaras se comieron a los perros chihuahueros

‘the tarahumaras SE-ate the chihuahuas’

(CDE)

▪ Non-edible/drinkable things

(114) allí está María, pegada a sus pies, bebiendo cada palabra y cada compás

‘there is Maria, stick to her foot, drinking every word and every beat’

(115) Jorge Lazarte, teme que ella termine comiéndose a la democracia

‘Jorge Lazarate is afraid the she ends up eating-SE the democracy’

(CDE)

For this factor I predict that unexpected objects, non-edible food or drinks (110)-(115) and (116)-(117), will favor SE-marking.

(116) ¿Bebió Malcolm After - Shave?

‘Did Malcom drink after-shave?’

(117) A Socorro se la comían los mosquitos

‘Socorro was SE being eaten by the mosquitos’

(CDE)

### 3.4.4 Subject associated factors

#### 3.4.4.1 Grammatical person and grammatical number

Two of the parameters associated with subject properties are *agency* and *volitionality*. The more animate the subject the more volitional it is. On the human scale of animacy, there is a differentiation between first person (primary speaker), second person (addressee) and third person (other):

It is natural that the speaker invests the strongest animacy/empathy for himself/herself, and the second strongest animacy/empathy in someone whom he/she is addressing in the course of a speech event (c.f. Langacker 1991). Another important point to note is that the use of first and second person deictic expressions entails direct reference to the speech participants (Yamamoto 1999: 3).

Subjects were coded for *grammatical person* (1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup>) and *grammatical number* (SG and PL) assuming that these two linguistic factors encode animacy and correspond with the *agentivity* and *volitionality* parameter in Hopper and Thompson's model. These parameters concern the degree of planned involvement of the agent in the activity of the verb: "Participants high in Agency can effect a transfer of an action in a way that those low in Agency cannot" (Hopper and Thompson 1980: 252). Volitionality refers to the presence of an agent acting purposefully. A participant will be taken to be volitional if its involvement in the event is defined by its being volitional or sentient. The effect is more evident if the agent acts purposefully. Regarding this parameter, it was

discussed in previous sections that volitionality and agency are closely related to the presence of SE, see examples (42)-(43), in which the use of SE yields pragmatically infelicitous sentences when the agent's volitionality is cancelled. Regarding *agency*, there is no question that participants high in agency can affect the transfer of an action in a way that those with low agency cannot.

Another methodological consideration for coding grammatical person was that previous studies have shown that first and second person subjects favor SE-marking and third person marking disfavors it in verbs like *subirse* and *bajarse* (Torres Cacoullos and Schwenter 2008).

#### 3.4.4.2 Subject animacy

I include animacy of the subject as a factor group in order to test two parameters associated with subjects: *agentivity* and *volitionality*. I operationalized these two parameters through animacy assuming that the more animate the subject, the more prototypical and the more agentive it will be. Therefore highly animate subjects (volitional subjects and agents high in potency) will be more likely to occur in highly transitive constructions. For animacy coding, I assume that animacy is not a single dichotomy between animate and inanimate, but rather “a continuum along which we can range entities according to their degree of animacy, so that for instance people are more animate than animals, and animals are more animate than inanimate objects” (Comrie 1989: 135). Within this factor group, I coded subjects for animacy in five categories, following the animacy hierarchy proposed by Croft (2003):

- *1<sup>st</sup> and 2<sup>nd</sup> person pronouns*
- *3<sup>rd</sup> person pronouns*
- *Proper names*
- *Human common nouns*
- *Non-human common nouns*

### **3.5 Statistical analysis**

Once the envelope of variation was delimited and the factor groups and factors were established, all tokens extracted from the CDE were analyzed using the variable rule program Goldvarb X (Sankoff, Tagliamonte and Smith 2005), the variable rule program not only allows for sophisticated statistical analysis but is also a very helpful tool that helps to make sense of linguistic data: “the variable rule was designed as an accountable, empirical model for this phenomenon, thus introducing a probabilistic component to the model of language” (Tagliamonte 2006: 130).

GoldvarbX is a statistical program, which identifies which factor groups that contribute a statistically significant effect on the dependent variable, in this case the presence of SE. In addition this program also shows the strength of each factor by means of the range. This program makes use of a type of multivariate analysis known as logistic regression and the program compares the expected distribution predicted by these estimates to the observed distribution and calculates the variance, or the distance between

the expected and observed distributions. As a measurement of variance, instead of a chi-square value, GoldVarb uses a value of log likelihood, a measurement of how well the model fits the data. To determine which factor groups exert a statistically significant effect, GoldVarb performs a step-up / step-down procedure. GoldVarb only performs binomial multivariate analysis, where there is a choice of two results: application vs. non-application. For each factor group, GoldVarb estimates the relative contribution (factor weight) that each factor in every factor group makes to the occurrence of the application value. In this case, the application value is the presence of SE marking.

### **3.6 Combining quantitative and qualitative data**

Besides all the great advantages that corpus-linguistic based studies present in terms of the amount of data, its variety and distribution, it is also true that it presents limitations: a specific construction can be missing in the corpus, or its occurrence may be too low to obtain statistically significant generalizations. For these reasons, I complemented the naturally occurring data obtained from the CDE with constructed examples, which allowed me to observe the subtleties not found in the corpus and to control for details in the discourse and context of utterance that are out of the linguist's control when working with corpus sources.

### 3.7 Conclusions

In this chapter I have outlined the theoretical and methodological approach of this dissertation, mainly a variationist linguistic investigation complemented with qualitative observations to better explain data not found or that occurs with low frequency in the corpus.

Table 6 shows a summary of the linguistic factors considered in this analysis associated with the Transitivity parameters proposed by Hopper and Thompsen (1980).

PARAMETERS	HIGH/LOW TRANSITIVITY	LINGUISTIC FACTOR
A. PARTICIPANTS	2 or more participants (A-O)-1 participant	N/A
B. KINESIS	Action/non-action	Clause type (declarative-interrogative)
C. ASPECT	Telic/atelic	Tense
D. PUNCTUALITY	Punctual/non-punctual	Tense
E. VOLITIONALITY	Volitional/non-volitional	Animacy, grammatical person, grammatical number
F. AFFIRMATION	Affirmative/negative	Polarity
G. MODE	Realis/irrealis	Indicative, non-indicative (subjunctive, conditional, imperative)
H. AGENCY	A high in potency/A low in potency	Animacy, grammatical person and grammatical number
I. AFFECTEDNESS OF O	O totally affected/O not affected	Object number, specificity and definiteness

J. INDIVIDUATION OF O	O highly individuated/O non individuated	Object number, specificity and definiteness
-----------------------	--	---

**Table 6. Linguistic factors associated with Transitivity (Hopper and Thompson 1980)**

For each Transitivity parameter proposed by Hopper and Thompson I assigned at least one linguistic factor, therefore the Transitivity model was operationalized and the Transitivity Hypothesis tested.

In this chapter I discussed the variationist approach to analyzing linguistic data as well as the essential assumptions of the linguistic variable. The variable nature of SE-marking requires both a theoretical perspective and methodological approaches that can account for the variation and that can possibly be extended to other verbs and will be helpful in the characterization of Spanish SE.

The use of Latin American and Peninsular dialects provides the diversity of varieties of Spanish to evaluate SE-marking across dialects. The period of time selected for this analysis, the 20<sup>th</sup> Century, allows us to account for the current distribution of the phenomenon, as a preliminary diachronic search of these forms have showed.

At the same time, the envelope of variation, i.e. the contexts where variation is possible, was circumscribed and the invariant, categorical cases where marking or non-marking occurred were ruled out and the procedure for extracting tokens was outlined. The inclusion and exclusion criteria were detailed and the procedure for the coding scheme was made explicit through a detailed description of each factor group and the

subsumed factors in each one making this study easily replicable.

## CHAPTER 4. RESULTS

In this chapter, I will present the results of the quantitative analysis of coded tokens extracted from the *Corpus del Español* (CDE). The overall distribution of the data is presented in section 4.1. In sections 4.2, 4.3 and 4.4, I will present and closely analyze the results of the different groups selected in the multivariate analysis for *comer*, *tomar* and *beber* respectively. In section 4.4 I also include a dialectal analysis of *beber*, which only for this verb came out as a significant factor. Section 4.5 will be devoted to the analysis of the effect of the progressive on SE marking for the three verbs. The qualitative analysis interpreted in light of the quantitative results discussed in previous sections will be presented in section 4.6. A summary of the chapter is presented in section 4.7. Further discussions of these results as well as their theoretical implications will be addressed in chapter 5.

### 4.1 Overall statistical analysis

The results presented in this chapter correspond to the analyses of a total of 3869 tokens. As was addressed in the previous chapter, these tokens include both marked and

non-marked forms of *comer*, *tomar* and *beber* found in transitive clauses and in contexts which variation could potentially occur, according to the envelope of variation defined in chapter 4.

Table 7 shows the overall distribution of the non-SE marked and SE-marked tokens for each verb. Out of all the tokens analyzed, 28% (1086) are marked with SE and 72% (2783) are not marked; the chi-square results show that the difference between the SE-marked forms and non-SE-marked forms is highly significant. This table also shows that the SE-marked form occurs 36% of the time with *comer*, 22% with *tomar* and 16% with *beber* and that the difference between the three verbs is significant.

	SE-MARKED	NON-SE-MARKED	TOTAL
<i>Comer</i>	36% (720)	64% (1301)	52% (2021)
<i>Tomar</i>	22% (252)	78% (867)	29% (1119)
<i>Beber</i>	16% (114)	84% (615)	19% (729)
TOTAL	28% (1086)	72% (2783)	100% (3869)
Chi-square = 130, p = 0.001 (df=2)			

**Table 7. Overall frequency of SE and non-SE-marked forms in Spanish ingestive verbs**

In the following tables I present the results from the multivariate analysis run with GoldVarb. As I mentioned before, this program estimates the relative contribution that each factor in every factor group makes to the occurrence of SE marking (application value). Results from the multivariate analysis are presented in tables with four columns.

The left-most column shows the *factor groups* (bold) that came out as significant for SE marking. The individual factor values for each group are presented immediately below. The factor groups are arranged vertically in decreasing order of magnitude based upon their ranges. Within each factor group the range is determined by calculating the difference between the highest and the lowest factor weights (probabilities). Ordering the factor groups by range captures the effect of a factor group in terms of the largest disparity between factors favoring and disfavoring SE marking. The second column, *probability*, lists the factor weights which are centered on the neutral value .5, such that anything above .5 favors application, while anything below .5 disfavors application. The relative position of factor weights vis-à-vis each other is what is relevant for interpreting results. The third column, *% SE-FORM* provides the percentages of the marked tokens per factor value (number of SE-marked tokens/total number of tokens) and the last column (overall data) shows the percentage of the data that each factor represents. The factor groups that failed to obtain statistical significance in the analyses are listed at the bottom of each table.

The results of the multivariate analysis are found in Table 8. This analysis required extensive examination of the data, and numerous reconfigurations of the factor groups were necessary in order to identify the most accurate means for its interpretation.

FACTOR GROUPS	PROBABILITY	% SE-FORM (SE/TOTAL N)	OVERALL DATA
<b>Verb Type</b>			
<i>Comer</i>	.56	36% (720/2021)	52%
<i>Tomar</i>	.52	22% (252/1119)	29%
<i>Beber</i>	.30	16% (114/729)	19%
	<b>Range 28</b>		
<b>Definiteness</b>			
Definite	.63	42% (722/1718)	44%
Indefinite	.39	17% (364/2151)	56%
	<b>Range 24</b>		
<b>Edibility/drinkability</b>			
Non-edible/non-drinkable	.69	48% (351/724)	19%
Edible/drinkable	.45	23% (735/3145)	81%
	<b>Range 24</b>		
<b>Animacy</b>			
Non-human common nouns	.67	53% (162/307)	8%
3 <sup>rd</sup> person pronouns, proper names	.52	29% (441/1554)	40%
1 <sup>st</sup> and 2 <sup>nd</sup> person pronouns	.46	25% (353/1402)	36%
Human common nouns	.43	24% (483/2018)	16%
	<b>Range 24</b>		
<b>Mood</b>			
Imperative, conditional	.71	45% (52/116)	3%
Subjunctive	.58	30% (61/204)	5%
Indicative	.49	27% (973/3549)	92%
	<b>Range 22</b>		
<b>Specificity</b>			
Specific	.55	34% (909/2668)	69%
Non-specific	.39	15% (177/1201)	31%
	<b>Range 16</b>		
<b>Register</b>			
Oral	.61	35% (382/1105)	29%
Written	.45	25% (704/2764)	71%
	<b>Range 16</b>		
<b>Aspect</b>			
Progressive	.61	37% (129/350)	9%
Perfectives	.55	32% (187/586)	15%
Present perfect	.51	31% (38/122)	3%
Imperfectives, other	.47	26% (732/2811)	73%
	<b>Range 14</b>		
<b>Polarity</b>			
Affirmative	.51	29% (1026/3570)	92%
Negative	.37	20% (60/299)	8%
	<b>Range 14</b>		
<b>Grammatical number (subject)</b>			
Singular	.54	30% (799/2655)	69%
Plural	.42	24% (287/1214)	31%
	<b>Range 12</b>		
<b>Object position</b>			
Preverbal	.56	39% (283/722)	81%
Post-verbal	.48	25% (803/3147)	19%
	<b>Range 6</b>		
Log likelihood -1921.359, Significance =.0005, Input= 0.237.			
Factor groups not selected as significant: Dialect, Grammatical person, Clause type, Grammatical number (object).			

**Table 8. Internal factors contributing to the choice of SE in Spanish for all verbs**

Two procedures were followed; factors with poor distributions in the data were collapsed with other factors within the same factor group, and factor groups that turned out to be non-significant were excluded for subsequent reanalysis. This recoding and exclusion procedure will be discussed when applicable in the following sections.

The factor group with the largest effect was verb type, for which *comer* and *tomar* favor the use of SE and *beber* disfavors it. The difference across verbs is statistically significant (chi square=75 and  $p=.001$ ). This result reveals that despite belonging to the same type of predicate i.e. verbs of ingestion, *comer*, *tomar* and *beber* belong each one of these verbs behaves different with respect to SE marking in Spanish.

Following the factor group verb type, the next factor groups that came out significant, both with the same range, were definiteness and edibility/drinkability. With respect to definiteness, the data reveal that definite objects strongly favor the occurrence of SE while indefinite objects strongly disfavor SE marking. This result was not surprising since definiteness is one of the most recurring factors constraining the use of aspectual SE mentioned in previous literature. Unlike definiteness, edibility/drinkability has not been mentioned in any of the previous literature as a possible constraint on SE variation and, as we will see later on, this factor came out significant in all the analyses, thereby suggesting that it is a crucial factor in SE marking. This overall analysis shows that non-edible/drinkable objects highly favor SE marking, while edible and drinkable objects slightly disfavor it.

The next factor group that came out significant was mood. The results for this factor group show that imperative and conditional forms highly favor the use of SE (.71), subjunctive forms favor it and indicative forms neither favor nor disfavor SE marking (.49). According to the Transitivity Hypothesis (SE being favored in highly transitive clauses), the prediction for this result was that indicative forms (realis) should favor marking and non-indicative forms (irrealis) should disfavor SE marking. Results for the overall analysis reveal the opposite, however this result can be explained through the counter-expectation hypothesis (CH). Indicative and non-indicative forms are paradigmatic alternatives (Croft 2003) and following Greenberg's basic markedness pattern for Mood (indicative < hypothetical) it is expected that hypothetical forms (non-indicative) are marked (Croft 2003). This result can also be interpreted through the counter-expectations hypothesis (CH); non-indicative forms correspond to the non-canonical mood (not expected) and assuming that one of the contributions of SE is to mark non-canonical events or grammatical categories non-indicative forms would favor SE marking.

Initially, subject animacy was coded for five factors that were collapsed in the last and best run in three groups, the explanation for this will be given in the following chapter. Surprisingly, the factor that most favors the use of SE was non-human common nouns followed by third person and proper names. Human common nouns and first and second person pronouns seem to have little effect on disfavoring marking. Animacy is another factor that has not been widely discussed in the previous literature on aspectual

SE, with the exception of variationist studies; the results suggest that animacy plays an important role in the variation of SE at least for verbs of ingestion. This result has implications for the counter-expectation hypothesis (CH). According to this hypothesis, SE would be likely to occur whenever there is a non-canonical event, subject or object. This result reveals that SE is highly favored when the grammatical subject is a non-human common noun, which corresponds to a non-canonical subject.

Specific objects slightly favor the use of SE but non-specific objects highly disfavor it. This pattern was already mentioned in the literature (although not discussed in depth as definiteness has been); objects strongly tend to be definite and specific when co-occurring with SE. This result goes along with the Transitivity Hypothesis (TH) that predicts that SE would be favored if the object of the predicate is highly individuated and can be totally affected (specific) and disfavored if the object is not affected and non-individuated (non-specific).

With respect to register, results reveal that SE marking is highly favored in oral register (.61) and slightly disfavored in written registers. Although, I did not make any prediction with respect to this factor group, this result reveals that SE marking is also discourse-dependent.

Following the predictions and hypotheses (AH and TH), aspect was one of the factor groups expected to be significant in SE marking, and it was. However it was also predicted that perfective aspect would highly favor the use of SE, assuming that SE is an aspectual marker, and this was not necessarily the case. Progressives rather than

perfectives tended to highly favor the use of SE, however it is still evident that SE contributes to a perfective interpretation. With respect to imperfectives, the results showed what was predicted; imperfects disfavored SE marking.

Regarding clause-verb associated factors, polarity gave interesting results. Affirmative clauses do not favor or disfavor marking (.51) but negative clauses highly disfavor SE marking (.37).<sup>11</sup> The Transitivity Hypothesis predicted that SE marking would be favored in affirmative clauses and disfavored in negative clauses. This prediction was true for negative clauses, which highly disfavored SE marking.

Grammatical number was the next factor group that came out significant. Singular subjects had a small effect favoring SE marking (.54), while plural subjects disfavored SE marking (.42). These results could also be interpreted within the transitivity framework as a factor that encodes agentivity. Assuming that plural subjects are “agents low in potency” and assuming that singular subjects are “agents high in potency” this result supports the TH; SE is a transitivity marker and its presence is associated with high transitivity features (A high in potency) while its absence is associated with low transitivity features (A low in potency).

The last factor group that was significant was object position. Pre-verbal objects favor SE and post-verbal objects disfavor it. The TH makes no prediction for object position however the CH could account for this result; objects in pre-verbal position do

---

<sup>11</sup> As in the case of non-edible/drinkable objects and animacy, this factor has been associated with counter-expectations in previous variationist studies, but it has only been studied in intransitive motion verbs

not correspond to the canonical word order of Spanish, therefore if SE were associated with non-canonical marking it would be predicted that SE would be favored in non-canonical word orders and it was.

It is also worth mentioning that *beber* strongly disfavors marking, and although this might be an effect of dialectal differences it is relevant to mention it, In the following sections I will discuss the dialectal and linguistic implications of this result.

When analyzing the type of factors that came out significant it is interesting to observe that out of the ten factors, four of them are either verb or clause related (verb type, mood, aspect and polarity), the same number for object-related factors (definiteness, edibility/drinkability, specificity and object position). Finally, only two subject-related factors turned out to be significant (animacy and grammatical number). Although these are raw results, they suggest that SE marking is not only determined by object properties, but also by subject and verb-clause associated properties, associated with grammar and pragmatics.

The last table (Table 9) summarizes the factor groups that came out significant for SE marking for all verbs. The numbers in the left column represent the order in which the factors were selected as significant, thus number 1 corresponds to the factor group with the largest effect on SE marking. The factor group with the largest effect was verb type. This result supports the argument for the lexically specific effects of SE since it reveals that verb type matters for the distribution of SE marking. With the exception of edibility/drinkability and register, all the factor groups selected as significant are

associated with Transitivity; at the same time the factors that favor SE marking are associated with High Transitivity.

	FACTOR GROUP	FACTOR ASSOCIATED
1	Verb type	Verb/clause
2	Definiteness	Object
2	Edibility/Drinkability	Object
2	Subject animacy	Subject
3	Mood	Verb/clause
4	Specificity	Object
5	Register	N/A
6	Aspect	Verb/clause
7	Polarity	Verb/clause
8	Grammatical number (subject)	Subject
9	Object position	Object

**Table 9. Summary of results for the overall analysis**

Since verb type was selected as a significant factor with the highest range, I decided to run independent analyses for each verb to further investigate the lexically specific effects that could be observed in the factors selected as significant for each verb. This analysis will serve to test or reject the hypothesis regarding the lexical contribution of each verb on SE marking. In the following sections I discuss the results obtained for each verb.

## 4.2 Statistical analysis for COMER(SE)

For the analysis of *comer(se)* a total of 2021 tokens were included. These tokens correspond to 52% of the data. 64% of the data correspond to non-marked tokens (1301) and 36% of the data is SE marked (720).

The results for the multivariate analysis of linguistic factors for *comer(se)* are presented in Table 10. A first look at the results shows that with the exception of mood and polarity, the same factor groups came out significant as in the overall analysis in Table 8, but are nevertheless differently ordered according to ranges. Just as in the overall results, the highest ranked factor group was definiteness. As these results show, definite objects strongly favor the use of SE with *comer* (.67) and indefinite objects highly disfavor it (.32), supporting previous claims about definiteness being crucial in these constructions. This result also supports the TH that predicted that highly definite objects would favor SE marking and indefinite objects would disfavor it.

The second highest ranked factor group was animacy. Initially, animacy was a factor group that included five factors. After many attempts to find the best distribution of the data it was found that the most significant difference and the best log likelihood was obtained by collapsing animacy into four factors; non-human common nouns, third person pronouns, proper names and 1<sup>st</sup> 2<sup>nd</sup> pronouns and human common nouns. Results reveal that same as in the overall analysis, SE marking is more likely to occur with non-human common nouns (.68) as the CH predicted.

FACTOR GROUPS	PROBABILITY	% SE-FORM (SE/TOTAL N)	OVERALL DATA
<b>Definiteness</b>			
Definite	.67	54% (556/1036)	51%
Indefinite	.32	17% (164/989)	49%
<b>Range 35</b>			
<b>Animacy</b>			
Non-human common nouns	.68	57% (159/279)	14%
Third person pronouns	.50	35% (244/704)	35%
Proper names	.47	33% (18/54)	3%
1 <sup>st</sup> and 2 <sup>nd</sup> pronouns, human dmis. Nouns	.44	30% (299/984)	49%
<b>Range 24</b>			
<b>Aspect</b>			
Progressive	.68	55% (88/160)	8%
Perfectives	.55	42% (111/354)	17%
Imperfectives	.47	32% (135/422)	21%
Other	.45	32% (352/1085)	54%
<b>Range 23</b>			
<b>Edibility</b>			
Non-edible/non-drinkable	.66	54% (256/474)	24%
Edible/drinkable	.45	30% (464/1547)	76%
<b>Range 21</b>			
<b>Mood</b>			
Imperative and conditional	.67	43% (31/72)	3%
Subjunctive	.57	35% (45/135)	7%
Indicative	.49	33% (644/1814)	90%
<b>Range 18</b>			
<b>Specificity</b>			
Specific	.55	43% (624/1453)	72%
Non-specific	.38	17% (96/568)	28%
<b>Range 17</b>			
<b>Register</b>			
Oral	.62	42% (256/616)	30%
Written	.45	33% (464/1405)	70%
<b>Range 17</b>			
<b>Object Position</b>			
Pre-verbal	.59	48% (247/518)	26%
Post-verbal	.47	31% (473/1503)	74%
<b>Range 12</b>			
<b>Polarity</b>			
Affirmative	.51	37% (673/1824)	90%
Negative	.41	24% (47/197)	10%
<b>Range 10</b>			
<b>Grammatical number (subject)</b>			
Singular	.53	38% (508/1352)	67%
Plural	.43	32% (212/669)	33%
<b>Range 10</b>			
Log likelihood: -1052.149, Significance: 0.045, Input .313			
Factors not selected as significant: Grammatical person, Clause type, Grammatical number (object), Dialect			

**Table 10. Internal factors contributing to the choice of SE in Spanish for *comer***

The third factor group that came out as significant was aspect.<sup>12</sup> For this factor group results progressive was the factor selected for having the strongest effect and showed a robust probability to favor SE (.68). The next factor selected in this factor group was preterit (perfective aspect), which slightly favors the application of the variable (.55). This result goes along with the predictions that aspectual approaches support; SE is more likely to occur with perfective predicates and SE is less likely to occur with imperfective predicates.

The fourth factor selected as significant was edibility/drinkability. This result follows the same pattern that was described in the previous section; non-edible/drinkable objects strongly favor marking and edible/drinkable objects disfavor marking. This result supports the CH that states that SE is a counter-expectation marker and it would be more likely to occur with non-canonical or counter-expected objects, in this case non-edible/drinkable objects.

The next factor that was selected as significant for SE marking with *comer* was mood. Same as in the overall analysis, the presence of SE is highly favored for non-indicative mood. Imperative and conditional (.67) and subjunctive (.57) highly favor SE marking.

Specificity appears in the sixth position of the table. Although specific objects do not have a robust effect in favoring SE (.55) non-specific objects have a strong

---

<sup>12</sup> For the analysis of all linguistic factors with the verb *comer*, present perfect and preterit both had the same weights therefore I collapsed the two factors in one factor group.

disfavoring effect on SE marking (.38). This result is another evidence for the TH, in which SE marking is associated with high transitivity features (individuation and affectedness of object).

Regarding object position, *comer(se)* results reveal that this form is more likely to occur with pre-verbal objects (.59). Object position was another factor selected as significant in SE marking for the verb *comer*; preverbal objects have a higher probability of being marked while post-verbal objects slightly disfavor marking.

For polarity, results show that SE is disfavored in negative clauses a feature associated with low transitivity and that supports the TH; SE marking is more likely to occur in affirmative clauses and less likely to occur in negative clauses.

The last factor that turned out to be significant was the grammatical number of the subject. Singular subjects are more likely to occur with SE than plural subjects, the former factor slightly favors (.53) and the latter strongly disfavors the variable marker (.43) another result that supports the Transitivity Hypothesis (TH).

In order to investigate interactions between factor groups I checked for significant changes in weights for all factors across the regression analysis as well as for gaps in cross-tabulations. Cross-tabs revealed some categorical data. Two of the interactions were normal by-products of the language, which means that they are not attested in Spanish, this means that the combination of for example progressive and subjunctive is not possible in Spanish. This was the case of progressive and subjunctive and preterit and imperative. Another interaction was found between progressive and imperative mood. In

Spanish it is possible to find examples of imperative-progressives like *andante, andante* ‘walk, walk’, but these uses are restricted to some verbs, and verbs of ingestion are among those, therefore I considered that this is also a case that is not attested in Spanish. Another gap in the data corresponded to inanimate subjects and first person. These cases are not expected in spoken Spanish but could be found in oral register *Yo, el árbol, me estoy comiendo el pasto* ‘I, the tree, is ME eating the grass’, although tokens with this referent were not found in this corpus.

In depth analysis and discussion of the results presented here in the light of theoretical models will be presented in the following chapter; nonetheless, a brief discussion of the most relevant findings can be advanced in this section.

First of all it is worth mentioning that when compared with the overall analysis of SE marking including all verbs, the factors that came out significant are almost the same, with the exception of polarity and mood which did not turn out to be significant for *comer*. All the factors that came out significant for *comer* were also significant for the overall analysis in Table 8. Besides the similarities in the factors selected in these analyses, there are interesting differences as well. In spite of the fact that the factors selected as significant for *comer* were also significant in the overall analysis, the ranking of these factors is not the same, which reveals potential differences across verbs of ingestion. Another striking result obtained from these analyses is that the majority of the factor groups selected as significant for *comer* are object related (definiteness, edibility, specificity, object position). Besides the proportion the type of factors selected show the

relevance of pragmatically and grammatically encoded information not considered in most of the previous literature. As in the overall analysis, the two factors related to grammatical subject were animacy and grammatical number. This pattern shows that subject characteristics also play an important role in SE variation, which might be associated with subjectivity as Aaron and Torres Cacoullos (2005) claim.

The biggest difference between the overall analysis and the analysis of *comer* is observed in the verb/clause related factors. In the case of *comer*, only aspect came out significant, as opposed to the overall analysis in which mood and polarity were also selected.

A summary of the factor groups selected as significant for *comer* is presented in Table 11.

FACTOR GROUP	FACTOR ASSOCIATED
1 Definiteness	Object
2 Animacy	Subject
3 Aspect	Verb/clause
4 Edibility	Object
5 Mood	Verb/clause
6 Specificity	Object
6 Register	N/A
8 Object position	Object
9 Polarity	Verb/clause
9 Grammatical number (subject)	Subject

**Table 11. Summary of results for *comer(se)***

### 4.3 Statistical analysis for TOMAR(SE)

Out of the 3869 tokens analyzed, 1119 (29%) corresponded to the verb *tomar*, with the variants distributed as follows; 867 tokens (78%) occurred without SE and 252 occurred with SE (22%).

After many regression runs and a close analysis of the distribution of the data, it was necessary to collapse factor groups. These decisions were made based on the weights (whether the factor favored marking or not) and also based on the theoretical implications of combining two factors within the same factor group.

With respect to animacy, the distribution of the data for this verb was quite different from that of other verbs. Because no tokens of *tomar* had a non-human common noun subject, only five factors are included for animacy with this verb, as opposed to six for the other verbs. A first regression run revealed that the main animacy distinction for this verb was not determined by an opposition between human *versus* non-human subjects but rather by a distinction of third person pronouns and proper names from the rest of the factors, therefore animacy was collapsed into three factors: 3<sup>rd</sup> person pronouns and proper names, 1<sup>st</sup> and 2<sup>nd</sup> person pronouns, and human common nouns.

Another factor group for which I collapsed factors was aspect. In the first run, preterit, progressive, imperfect and other comprised the factors in the factor group. However, the regression analysis revealed no effect for imperfect and other, thus I collapsed these two groups and as a result the log likelihood was improved.

A factor that seemed to be problematic in all runs was mood; there was great fluctuation in the data and radical changes both in the weights and in the arrangement of the factors. In order to investigate whether there were other factors that had a poor distribution or that were interacting with each other I tested for interactions.

In this case two different strategies were followed to test for interactions; first of all I checked in the step-up/step-down regression results for changes in the factor weights from one factor to another. It is assumed that when these changes are small and they do not affect the way in which factor effects are ordered by size then we may generally attribute those changes to sampling fluctuation (Tagliamonte 2006: 151). The other strategy was to cross-tabulate every factor group with every other factor group and observe the gaps in the data. After checking for cross-tabulations across the data, I went through the regression results to check which of the factor weights changed dramatically at the point affecting the way they are ordered. This analysis revealed that the most significant interaction occurred with imperative and subjunctive, which happened to represent a low percentage of the data (3% each). Because the two factors favored the use of SE, I collapsed the two factors (subjunctive and imperative) in one factor. The regression analysis after collapsing these factors, improved the log likelihood from -566.544 to -564.618 and the  $p$  value remained the same (.009).

The results obtained in the final regression analysis are presented in Table 12.

FACTOR GROUPS	PROBABILITY	% SE-FORM (SE/TOTAL N)	OVERALL DATA
<b>Edibility/drinkability</b>			
Non-edible/non-drinkable	.77	48% (50/105)	9%
Edible/drinkable	.47	20% (202/1014)	91%
	<b>Range 30</b>		
<b>Animacy</b>			
3 <sup>rd</sup> person pronoun, proper names	.56	26% (110/427)	38%
Non-human common nouns	.49	14% (16/116)	10%
1 <sup>st</sup> and 2 <sup>nd</sup> person pronouns	.32	22% (126/576)	52%
	<b>Range 24</b>		
<b>Polarity</b>			
Affirmative	.51	23% (244/1061)	95%
Negative	.32	14% (8/58)	5%
	<b>Range 19</b>		
<b>Mood</b>			
Non-indicative	.66	34% (25/73)	7%
Indicative	.49	22% (227/1046)	93%
	<b>Range 17</b>		
<b>Specificity</b>			
Specific	.55	27% (178/667)	60%
Non-specific	.42	16% (74/452)	40%
	<b>Range 13</b>		
<b>Register</b>			
Oral	.56	26% (101/394)	35%
Written	.47	21% (151/725)	65%
	<b>Range 9</b>		
Log likelihood: -559.799 Significance: 0.044, Input 0.209			
Factors not selected as significant: Grammatical person, Grammatical number (subject), Aspect, Definiteness, Clause type, Grammatical number (object), Object position, Dialect			

**Table 12. Internal factors contributing to the choice of SE in Spanish for *tomar***

Out of eleven factors six were selected as significant. It is interesting to observe that all the factor groups selected as significant for *tomar* coincide with previous results for the overall analysis and for *comer*. Unlike the overall analysis and the *comer* analysis, mood was selected for this verb as a significant factor group.

Surprisingly definiteness was not selected as significant. Instead the factor group with the highest range was edibility/drinkability. The same pattern as in previous results was observed for this factor group; non-edible/drinkable objects highly favor SE marking (.77) and edible objects slightly disfavor SE marking (.47).

Animacy was the second factor selected as significant. Unlike the case of *comer*, in which the primary distinction was found between human and non-human subjects, for *tomar* the distinction is more specific. Within this factor group it matters whether the subject has a third person as a referent (third person pronouns and proper names), if it is a 1<sup>st</sup> or 2<sup>nd</sup> pronoun or a non-human common noun. Third person and proper names favor SE marking (.56), while 1<sup>st</sup> and 2<sup>nd</sup> person pronouns highly disfavor SE marking (.32).

The third factor that came out as significant was polarity. This results supports the TH; negative clauses highly disfavor SE marking (.32).

The next factor group that turned out to be significant was mood. After many runs and close analysis of the data distribution, the best arrangement of the data was to collapse subjunctive, imperative and conditional, reported as non-indicative. For this factor group, the factor that has the strongest effect on SE marking is non-indicative which strongly favors the use of SE (.66), while indicative neither favors nor disfavors marking (.49).

Specificity was the fifth factor selected as significant for SE marking with *tomar*. Following the same pattern as seen for *comer* and in the overall analysis for all verbs. Specific objects favor the use of SE while non-specific objects disfavor it.

It is worth remembering that even though the most frequent use of *tomar* is related to the event of drinking (118), it is also possible to use it with a sense of *eating* (119) and it is possible that due to this semantic flexibility the behavior of this verb shares some marking properties with both *comer* and *beber*. These semantic similitudes are reflected in the results.

(118) Me tomé una cerveza

‘I ME drank a beer’

(119) Me comi una hamburguesa

‘I ME ate a hamburger’

With respect to the factor groups selected as significant with the verb *tomar* it is interesting to observe that out of four factors, two were object-related (non-edible/drinkable and specificity), one verb-clause related (mood) and one subject-related (animacy). Three of them are referential categories (non-edible/drinkable object, specificity and animacy) and are semantic or pragmatic by nature. On the other hand, with the exception of definiteness, the linguistic factors that were not selected as significant all encode grammatical categories, overtly marked in the language (i.e. grammatical person, grammatical number (subject), aspect, clause type, etc.).

Table 13 presents a summary of the results for variable use of SE with *tomar*. Same as previous results, this table shows that variable SE marking is constrained by subject, object and clause/verb associated factors.

FACTOR GROUP		FACTOR ASSOCIATED
1	Edibility/drinkability	Object
2	Animacy	Subject
3	Polarity	Verb/clause
4	Mood	Verb/clause
5	Specificity	Object
6	Register	N/A

**Table 13. Summary of results for *tomar(se)***

#### **4.4 Statistical analysis for BEBER(SE)**

Because of the reasons discussed in the previous chapter, the verb *beber* had a special treatment. A first regression analysis revealed that dialect was a significant factor in SE marking, because of this reason and due to the dialectal differences that were discussed in previous chapter related to this verb, besides the separate verb analysis of *beber*, I also ran one regression analysis for Latin American Spanish and another for Peninsular Spanish. The overall results of *beber* and the results according to dialect are presented in the following paragraphs.

Table 14 shows the overall distribution of SE-marked and non-SE-marked forms in *beber* across Latin American and Peninsular Spanish. The chi square test reveals that there is no significant difference in marking SE across Latin American and Peninsular varieties.

	SE-MARKED	NON-SE-MARKED	TOTAL
Latin American Spanish	17% (82)	83% (397)	66% (479)
Peninsular Spanish	13% (32)	87% (218)	34% (250)
TOTAL	16% (114)	84% (615)	100% (729)
Chi-square = 2.32, p = 0.128 (df=1), df=1			

**Table 14. Distribution of *beber(se)* in Latin American and Peninsular Spanish**

In order to have parallel results to those discussed in previous sections, I first ran a regression analysis for internal factors including both dialects. Because of the dialectal differences discussed previously, most of the analysis will derive from the results obtained per dialect, however it was necessary to be consistent with the analysis, therefore the analysis for *beber* with two dialects is also included in this section.

In the first regression analysis a total of 729 tokens were included (19% of the data); 114 of these tokens (16%) were SE marked and 615 (84%) were unmarked. After many regression runs and a close analysis of data distribution and weights, the best run required the collapsing of imperative and conditional in one group. Although these factors only encompass 1% of the data (9 tokens), its exclusion resulted in a worse log likelihood and *p* value, for this reason the factor was maintained for the analysis. Multivariate analysis results are presented in Table 15.

Mood turned out to be the factor group with the largest effect. Non-indicative forms strongly favor SE marking (.92) and indicative forms slightly disfavor SE (.48).

FACTOR GROUPS	PROBABILITY	% SE-FORM (SE/TOTAL N)	OVERALL DATA
<b>Mood</b>			
Imperative, conditional	.94	67% (6/9)	1%
Subjunctive	.64	19% (6/31)	4%
Indicative	.48	15% (102/689)	95%
	<b>Range 46</b>		
<b>Specificity</b>			
Specific	.59	19% (107/548)	75%
Non-specific	.24	4% (7/181)	25%
	<b>Range 35</b>		
<b>Polarity</b>			
Affirmative	.52	16% (109/685)	94%
Negative	.23	11% (5/44)	6%
	<b>Range 29</b>		
<b>Register</b>			
Oral	.74	26% (25/95)	13%
Written	.46	14 (89/64)	87%
	<b>Range 28</b>		
<b>Grammatical number (subject)</b>			
Singular	.57	19% (102/542)	74%
Plural	.30	6% (12/187)	26%
	<b>Range 27</b>		
<b>Edibility/drinkability</b>			
Non-edible/non-drinkable	.71	31% (45/145)	20%
Edible/drinkable	.44	15% (103/691)	80%
	<b>Range 27</b>		
<b>Definiteness</b>			
Definite	.60	24% (82/342)	47%
Indefinite	.40	8% (32/387)	53%
	<b>Range 20</b>		
<b>Dialect</b>			
Latin American Spanish	.55	17% (82/479)	66%
Peninsular Spanish	.41	13% (32/250)	34%
	<b>Range 14</b>		
Log likelihood: -257.417, Significance= 0.034, Input .101 Factors not selected as significant: Register, Object position, Grammatical number (object), Grammatical person, Animacy, Clause Type, Aspect,			

**Table 15. Internal factors contributing to the choice of SE in Spanish for *beber***

Specificity was the second factor group selected as significant. Results were the pattern of the results was exactly the same as in previous analyses: specific objects strongly favor the use of SE (.60) and non-specific objects strongly disfavor it (.19). Results show that specific objects favor marking and non-specific objects highly disfavor it, results that support the TH.

The results associated with the polarity parameter confirm the TH; negative clauses highly disfavor SE marking (.23). For register it was found that oral register highly favors the use of SE with *beber* (.74) and written register slightly disfavors it (.46).

The next factor group that came out significant was grammatical number of the subject. The results from this factor group reveal that singular subjects favor marking (.58) while plural subjects highly disfavor marking (.28). The same pattern was also found and discussed for *comer*, *tomar* and in the overall analyses.

Another pattern that was repeated and that supports the CH was the role of counter-expectations on SE marking, operationalized through edibility/drinkability. For *beber(se)* I found that non-edible/drinkable objects highly favor the use of SE (.72) and edible/drinkable objects slightly disfavor it (.44).

The penultimate factor selected as significant was definiteness. The results show that same as previous analyses; definite objects highly favor SE (.60) and indefinite objects highly disfavor it (.40).

The last factor that came out as significant was dialect, for this reason I decided to run an independent analysis for each dialect to test whether the factors that were

constraining the variation were the same for Latin American Spanish and for Peninsular Spanish. A summary of results for analysis of *beber (se)* is presented in Table 16.

	FACTOR GROUP	FACTOR ASSOCIATED
1	Mood	Verb/clause
2	Specificity	Object
3	Polarity	Verb/clause
4	Register	N/A
5	Grammatical number	Subject
5	Edibility/drinkability	Object
6	Definiteness	Objects
7	Dialect	Verb/clause

**Table 16. Summary of results for *beber(se)***

In the following section, I discuss these findings in depth in light of the results across dialects.

#### **4.5 Statistical analysis of BEBER(SE) across dialects**

As I discussed in Chapter 2, there are some clear differences in the use of *beber* and *tomar* across Spanish dialects; *tomar* is more common in Latin American Spanish and *beber* is more common in Peninsular Spanish. In Latin America, both *beber* and *tomar* could be used to refer to an exaggerate ingestion of alcoholic beverages although *tomar* is preferred for most of the dialects, while in Peninsular Spanish *beber* is the preferred form. Despite the fact that, in the first analysis, the results revealed that *beber* disfavors SE marking, I decided to run an independent analysis for two reasons: first, to see whether these results would confirm that there was indeed a dialectal difference in the

use of these verbs; and second, to further investigate the factors and hierarchy of factors that conditioned SE variation across and within verbs of ingestion in Spanish. Multivariate analysis results for *beber* reflected the dialectal difference discussed above since dialect was a significant factor. The analysis reveals that Latin American varieties favor the use of SE with *beber* and Peninsular Spanish disfavors it. The next step then was to investigate whether SE marking was constrained by the same factors across dialects.

#### **4.5.1 BEBER(SE) in Latin American Spanish**

For the analysis of Latin American Spanish, I analyzed 479 tokens (12% of the data); out of the 479 tokens, 82 (17%) were marked and 397 (83%) unmarked.<sup>13</sup>

Imperatives, conditional and subjunctive showed very similar weights therefore they were collapsed in one group. The factor group in the analysis presented here includes two groups, indicative and non-indicative (subjunctive, imperative and conditional). Table 16 shows the results for *beber* in Latin American Spanish.

Five factor groups were selected as significant for SE marking of *beber* in this variety: mood, specificity, grammatical number, edibility/drinkability and definiteness. Six factor groups were not selected as significant: polarity, clause type, aspect, grammatical number (object), grammatical person and animacy.

---

<sup>13</sup> The factor group clause type was excluded from the analysis because no variation was observed in interrogative clauses; the verb only occurred in the unmarked form in such contexts

The results show that the same factors selected as significant for the overall analysis of the verb were also significant for Latin American dialects and the factor weights are arranged in the same order for the first two factor groups (mood and specificity).

FACTOR GROUPS	PROBABILITY	% SE-FORM (SE/TOTAL N)	OVERALL DATA
<b>Mood</b>			
Non-indicative	.92	56%(9/16)	3%
Indicative	.48	16% (73/463)	97%
	<b>Range 44</b>		
<b>Specificity</b>			
Specific	.60	21% (78/370)	77%
Non-specific	.19	4% (4/109)	23%
	<b>Range 41</b>		
<b>Grammatical number (subject)</b>			
Singular	.58	21% (75/361)	75%
Plural	.28	6% (7/118)	25%
	<b>Range 30</b>		
<b>Edibility/drinkability</b>			
Non-edible/non-drinkable	.72	35% (33/95)	20%
Edible/drinkable	.44	13% (49/384)	80%
	<b>Range 28</b>		
<b>Definiteness</b>			
Definite	.63	28% (60/216)	45%
Indefinite	.39	8% (22/263)	55%
	<b>Range 24</b>		
Log likelihood: -175.965, Significance: 0.001, Input: 0.109 Factors not selected as significant: Register, Polarity, Clause Type, Aspect, Grammatical number (object), Grammatical person, Animacy.			

**Table 17. Internal factors contributing to the choice of SE in Spanish for *beber* in Latin American Spanish**

In spite of the poor distribution of the data for mood, a chi-square test revealed that the difference found between non-indicative and indicative is significant (chi-square = 9.59 degrees of freedom = 1,  $p = 0.002$ ). This factor group has the highest range, which means that is the factor group with the largest effect on SE marking. Indicative mood slightly disfavors the application of the value (.48) compared with non-indicative forms, which highly favor SE marking (.92).

Specificity also constrains SE marking of *beber* in Latin American Spanish; specific objects highly favor marking whereas non-specific objects highly disfavor marking. The third factor group that came out as significant was grammatical number. For this factor group, results show that singular subjects favor SE and plural subjects highly disfavor it.

The analysis of *beber(se)* in Latin American Spanish also revealed that singular subjects are more likely to occur with SE (.58) and plural objects are less likely to occur with SE (.28). As I have discussed in previous analyses, these result support the TH.

Edibility/drinkability was the next factor group selected as significant. For this factor group, the results reveal that non-edible/drinkable objects highly favor SE marking (.72) whereas edible/drinkable objects slightly disfavor it (.44).

The last factor group that tuned out significant was definiteness. The same pattern observed for other verbs is observed for *beber* in Latin American Spanish; definite objects highly favor SE (.63) and indefinite objects highly disfavor it (.39).

When considering the factor groups associated with verb/clause, subject and object, we find the same pattern discussed for other verbs. The most frequent factor groups are object related (specificity, edibility and definiteness) followed by verb/clause (mood) and subject related factors (grammatical person). It is also interesting to observe that out of the factors that did not come out as significant three of them are verb/clause related (polarity, clause type and aspect), two of them subject-related (grammatical person and animacy) and one object-related (grammatical number of the object). As I have mentioned before I will discuss the theoretical and linguistic implications of these findings in the following chapter.

Because of the methodological limitations that the quantitative analysis of this verb has in terms of its distribution across dialects, I decided to add a brief description of the most interesting findings that a close analysis of the data shows. One of these findings is the distribution of grammatical person. In spite of not coming out as a significant factor, it has an interesting distribution. In Latin American Spanish the most frequent grammatical person is third person (81%), followed by first (17%) and second person (2%). Another interesting finding shows that most of the marked occurrences of *beber* in Latin America are associated with alcohol (half bottle of wine, my rum, a bottle of champagne) or with non-drinkable objects, either for the amount or for its properties (the juice of 50 oranges, to drink all your salary in vodka, blood) and a third category that was not coded in the corpus was the manner of ‘drinking’ for which complements such as *de un solo tragico* ‘in one gulp’ are very frequent:

- (120) El Charifas se había bebido *media botella de güisqui* y era incapaz de recordar  
 ‘Caritas’ SE had drunk *half a bottle of whiskey* and was incapable of remembering’
- (121) la madre, harta de que su marido se beba *el sueldo* en vodka  
 ‘the mother, sick of her husband SE spending (drinking) his *salary* on (in) vodka’
- (122) él pretendía beberse su sangre  
 ‘he tied to drink-SE *her blood*’
- (123) Simplemente se bebía una ginebra *de un trago*,  
 ‘(he/she) simply SE-drunk a gin *in one gulp*’

(CDE)

Although the discussion of this data will be addressed in the next chapter, it is possible that the verb *beber* is being used in Latin American Spanish as the marked verbal form for drinking and *tomar* is the unmarked form.

#### 4.5.2 BEBER(SE) in Peninsular Spanish

Out of the 729 tokens of the verb *beber*, 250 of them (34%), corresponded to Peninsular Spanish, distributed as follows: 218 tokens (87%) are unmarked and 32 (13%) are marked with SE. From the first run, the results show that there were some knockouts and singletons. In the case of the factor group animacy, non-human subjects were excluded from the analysis because only the unmarked form was found. In the case of proper names the same thing happened. Therefore I collapsed them with 3<sup>rd</sup> person pronouns. This run also revealed that perfectives and progressives strongly disfavor the application of the value with very similar weights, thus I also collapsed them.

Table 18 presents the regression analysis result of *beber* for Peninsular Spanish. Out of the eleven factor groups, four of them came out significant in the following order: animacy, aspect, specificity and edibility/drinkability. Factor groups that did not come out as significant are: mood, polarity, clause type, grammatical number (object), definiteness, grammatical person, grammatical number (subject).

FACTOR GROUPS	PROBABILITY	% SE-FORM (SE/TOTAL N)	OVERALL DATA
<b>Register</b>			
Oral	.86	29% (16/22)	22%
Written	.37	8% (16/194)	78%
	<b>Range 49</b>		
<b>Subject Animacy</b>			
Third person pronouns, proper names	.79	24% (18/76)	32%
Human common nouns	.37	7% (4/60)	25%
1 <sup>st</sup> and 2 <sup>nd</sup> person pronouns	.34	10% (10/103)	43%
	<b>Range 45</b>		
<b>Edibility/drinkability</b>			
Non-edible/non-drinkable	.67	24% (12/50)	20%
Edible/drinkable	.46	10% (20/200)	80%
	<b>Range 21</b>		
<b>Specificity</b>			
Specific	.61	16% (29/178)	71%
Non-specific	.25	4% (3/72)	29%
	<b>Range 36</b>		
Log likelihood: -651.902, Significance: 0.017, Input .253			
Factors not selected as significant: Polarity, Clause Type, Aspect, Grammatical number (object), Grammatical person, Animacy.			

**Table 18. Internal factors contributing to the choice of SE in Spanish for *beber* in Peninsular Spanish**

Animacy turned out to be the second highest ranked factor group based on the range. On the other hand, animacy was also significant for *tomar*, so it seems that the

marking of SE in *beber* in Peninsular Spanish is more like *tomar*. Another explanation could be that the ‘weird’ distribution and selection of factor groups for this verb occurs in the Latin American variety in which *beber* seems to be the marked form to refer to a drinking event. Within this factor group, 3<sup>rd</sup> person pronouns and proper names highly favor SE marking, 1<sup>st</sup> and 2<sup>nd</sup> person pronouns slightly disfavor marking and finally, human common nouns highly disfavor it.

Another interesting result is observed in the second factor group: aspect. Besides *beber* in Peninsular Spanish, this factor group only resulted significant for *comer* and in the overall analysis. For this group, perfectives and progressives highly disfavor SE marking, just as imperfectives and other forms highly favor marking.

For specificity and edibility *beber* in Peninsular Spanish follows the same pattern as found in the previous analyses of the other verbs. For specificity, the results show that SE marking is highly favored with specific objects and highly disfavored with non-specific subjects and finally for edibility, SE is more likely to occur with non-drinkable objects and it is less likely to occur with drinkable objects.

#### **4.6 Qualitative analysis. Progressive and counter-expectation contexts**

As I mentioned in the previous section, progressive tokens were coded without considering the tense/aspect of the auxiliary verb. In order to test if the aspect of the auxiliary verb played a role in the distribution and marking of SE in progressives, I

isolated all progressive tokens and recoded them based on the auxiliary verb aspect if there was any. After running a chi square, the difference was tested to be no significant (chi square=-.183, p=0.980), which means that the auxiliary verb does not play a significant role in SE marking, instead the marking it is determined by the aspect of the main verb, in this case progressive. Progressives comprised 9% of the data analyzed. Out of 350 tokens, 221 (63%) were unmarked and 129 (37%) marked.

The preference for progressive was one of the most striking and unexpected results. As has been discussed in the literature, SE is a marker of completeness. But if this is so, why would it be the case that SE marking is favored in progressives and that perfective predicates do not necessarily favor its occurrence? In order to answer this question I analyzed some of the progressive tokens to determine whether it was really aspect or something else that was determining the high occurrences of SE.

The overall analysis reveals that for the linguistic factor group aspect, progressive was the factor that favors SE the most (.61) followed by perfective aspect (.55). The same pattern, was observed for *comer*; progressives highly favor SE (.68) and perfectives slightly favor SE (.55).

Through the qualitative analysis of the progressive tokens, I found that *comerse*, *tomarse* and *beberse* occur in the following contexts: (i) eat/drink in a non-canonical manner, against social rules; (ii) eat/drink something not edible/drinkable or something edible and drinkable in exaggerate amounts; (iii) the ingestive event is performed by a non-human agent; (iv) eat/drink something up; (v) eat/drink to have a good time. With

the exception of (v), all these contexts make reference to an event that occurs against the normal expectations or desires of the speaker.

- (i) *Eat/drink in a non-canonical manner, against social rules (like eating someone else's food, eating the soup with a spoon, drinking a drink in one gulp etc).*

Example (a) illustrates an ingestion event that occurs against social rules, in this case, eating someone else's food: what triggers SE marking is the fact that someone else is eating the speaker's food. Example (b), shows an example in which the speaker refers to how fast they were eating the food and in (c), in which the event refers to a drinkable object, 'a drink', the presence of SE highlights the manner of drinking it, in this case in one gulp and finally (d) illustrates the speaker's point of view about people who is against hunting but still eat hunting animals:

- (a) Salgan, bastante trabajo tengo sin ustedes comiéndose *mi comida*. No los quiero ver  
'Get out, I have had enough with you eating-SE mi food. I don't want to see you all'
- (b) nos los íbamos comiendo *corriendo*, para que no se dieran cuenta de nuestra ausencia  
we NOS were eating them *very fast*, so they did not notice our absence'
- (c) ¡Qué me contás! - - dijo Traveler, bebiéndose una caña *de un trago*  
'What's new? – said Traveler, drinking-SE a drink *in one gulp*'
- (d) seguramente quienes critican que la caza es cruel, casi seguro, están comiéndose una perdiz estofada. Yo admitiría esa crítica de un vegetariano. Sería perfectamente admisible

‘I am sure that people that is against hunting, almost sure, they are eating-SE a roasted partridge. I would admit that critic from a vegetarian. That would be perfectly admissible’

(CDE)

- (ii) *Eat/drink something judged as the speaker as non-edible/drinkable and/or exaggerated amounts of food/liquids.*

In (e) and (f), it seems that SE is not highlighting the ongoing event of eating but the fact that the speaker is referring to a non-edible object, a fat dog and the democracy respectively. In (g), SE highlights the exaggerate amount of cocoa that rats have been eating.

- (e) Se imagina usted comiéndose *un perro gordito* bien asado así como lo sirven en los países asiáticos

‘Can you see yourself eating-SE *a fat grilled dog*, as it is served in the Asian countries?’

- (f) vocal de la Corte Nacional Electoral, Jorge Lazarte, teme que ella termine comiéndose *a la democracia*

‘representative of the National Elections Court, Jorge Lazarte, fears that she will end up eating-SE *the democracy*’

- (g) de mil millones de pesos al año los daños causados. Las ratas se están comiendo *unas ocho mil toneladas de cacao al año*, que es más de lo que ‘about a thousand million pesos for damages. The rats are SE eating about *eight thousand tons of cocoa* every year, which is more than we’

(CDE)

(iii) *The ingestive event is performed by a non-human agent*

- (h) El negocio de la droga se está comiendo los fundamentos de Colombia como nunca antes

Lit. 'The drugs business is SE eating Colombian principles like never before'

- (i) el maíz y los frijoles que guardaban en el tabanco, se los estaba comiendo el gorgojo

'The corn and the beans that they keep in the bodega were SE being eaten by the weevil'

- (j) Comadre, la hierba se está comiendo al trigo

'Comadre, the grass is SE eating the wheat'

(CDE)

(iv) *Eat/drink something up*

- (k) observó Polo Sánchez bebiéndose el coctail de un trago

'said Polo, drinking-SE the coctail up in one gulp;

(CDE)

(v) *Eat/drink to have a good time*

- (l) Y nada que es que estamos aquí tomándonos unas copas

'And we are here, having-NOS some drinks'

- (m) Le llegó a dar risa pensar que Joaquín estaba tan cerca, tomándose plácidamente su whisky

'He started laughing thinking about Joaquín drinking-SE his whisky'

(CDE)

## 4.7 Conclusions

In this chapter I have presented the quantitative results for multiple multivariate analyses of SE marking in *comer*, *beber* and *tomar*. Table 19 shows a summary of the findings discussed in this chapter. This table shows that the factors that came out significant for the *overall* analysis and for *comer* are exactly the same with different arrangement and slightly the same as *tomar*. For the overall and *comer* results, the factor groups are arranged in different order, on this basis it can be assumed that the factors selected by the group of verbs (ingestion) and the hierarchy of them is determined by the specific kind of verb. In the case of *tomar*, all the factors that came out significant for this verb were also significant for *comer* and for the overall analysis, with the exception of definiteness, aspect, object position and grammatical person. Although I will discuss this data in detail in the following chapter, it is important to remember that both, definiteness and aspect have been two of the most mentioned factors on SE literature.

Edibility and animacy of the subject, both associated with counter-expectations, were ranked higher in the three analyses. For all verbs they all occur within the first four places. Polarity came out as a significant factor group favored in all cases by affirmative clauses, as it was predicted by the transitivity hypothesis (TH). For the three verbs, the distinction between indicative and non-indicative is marked in the same way. For all verbs indicative has no effect on marking (.49), but non-indicative forms highly favor SE.

Specificity was one of the factors that were predicted to be significant for SE marking and this was the case. In all verbs, results show that specific objects favor the use of SE, another argument for the Transitivity Hypothesis (TH).

Some unexpected results were also found such as the prominence of the progressive in SE marking, a factor that has not been mentioned in previous literature and that will be addressed in the following chapter.

With the exception of edibility/drinkability and register, the three factors with the largest effect for SE marking overall and per verb are associated with transitivity, and the factors that favored SE marking correspond to high transitivity as it was predicted in the TH.

These results show what I argued in previous chapters; SE variation is not only constrained by aspectual factors but also by lexical, referential, pragmatic and discourse factors. They also reveal that the lexical properties of the verb also contribute to the factors that favored the use of SE in Spanish ingestive verbs.

Table 19 shows that the results presented in this dissertation support the Transitivity and Aspectual Hypothesis and also that within verbs of ingestion, specific properties of the verb matter for the variation. Out of the eleven factors selected as significant for SE marking in Spanish ingestive verbs, nine were related to the Transitivity Hypothesis and the factors that favor SE marking are all associated with high transitivity. Aspect was also selected as significant, which confirms that SE plays an important role marking perfective predicates. This factor is also associated with the

lexical properties of the verb; as Table 19 shows, Aspect only came out as a significant factor group for the overall analysis and for the analysis of *comer*, the fact that it did not come out as significant with verbs associated with the ingestion of liquids, shows that there is an interaction between SE marking of aspectual predicates and lexical properties of the verb. This Table also shows that the factors that came out as significant for all verbs are pragmatically and grammatically determined. Mood and polarity are grammatically determined and also associated with Transitivity. Specificity and edibility are pragmatically determined; specificity is a referential property that is also associated with Transitivity and edibility is associated with counter-expectations.

	ALL VERBS	COMER(SE)	TOMAR(SE)	BEBER(SE)
0	Verb type	N/A*	N/A	N/A
1	Definiteness	<b>1</b>	N/S**	6
1	<b>Edibility/drinkability</b>	4	<b>1</b>	5
1	Animacy	<b>2</b>	<b>2</b>	N/S
2	<b>Mood</b>	5	4	<b>1</b>
3	<b>Specificity</b>	6	5	<b>2</b>
4	<b>Register</b>	6	6	<b>3</b>
5	Aspect	<b>3</b>	N/S	N/S
6	<b>Polarity</b>	8	<b>3</b>	3
7	Grammatical number	9	N/S	4
8	Object position	7	N/S	N/S

Table 19. Factor groups selected as significant for all verbs, *comer(se)*, *tomar(se)* and *beber(se)*

\* Not applicable, \*\*Not significant

<b>BEBER(SE) BOTH DIALECTS</b>	<b>BEBER(SE) LATIN AMERICAN SPANISH</b>	<b>BEBER(SE) PENINSULAR SPANISH</b>
<b>1</b> Mood	<b>1</b>	N/S
<b>2</b> <b>Specificity</b>	<b>2</b>	4
<b>3</b> Polarity	N/S	N/S
4 Register	N/S	<b>1</b>
5 Grammatical number	<b>3</b>	N/S
5 <b>Edibility</b>	4	<b>3</b>
6 Definiteness	5	N/S
7 Dialect	N/A	N/S
N/S Animacy	N/S	<b>2</b>

**Table 20. Factor groups selected as significant for *beber(se)***

In summary these results show that SE marking is constrained by grammatical and pragmatic factors associated with Transitivity and counter-expectations respectively.

In the final chapter I will discuss the theoretical implications of these results as well as the conclusions.

## **CHAPTER 5. DISCUSSION AND CONCLUSIONS**

In this chapter I will discuss the theoretical implications of the results presented in the previous chapter and based on the hypotheses raised in Chapters 1 and 2. In section 5.1 I will discuss the results of the analysis in relation to the hypothesis that SE is a transitivity marker and its presence correlates with high transitivity. In this section I also discuss the role of SE as an aspectual marker, not as an isolated factor but as part of the transitivity model. Section 5.2 is devoted to the analysis of SE as a counter-expectation marker associated with subjectivity. In this section, I will also discuss what these results tell us with respect to the hypothesis that SE highlights verb properties, and therefore the factors that constrain variation will be to some extent verb-dependent. In section 5.3, I discuss how the results presented here show that SE variation is constrained by pragmatic and grammatical factors, and finally, I will discuss why SE can be considered a marker of non-canonical events.

Section 5.4 will be devoted to the discussion of the extension of the Transitivity model to the analysis of variable SE in Spanish with different verb-types. A re-analysis of

the Transitivity model proposed by Hopper and Thompson (1980) is presented in section 5.5 and the conclusions of this work will be presented in 5.6. In the last section, I discuss the future research that can be derived from the present work.

As a result of the literature review and also due to the lack of agreement and problematic theoretical approaches to the uses of the so-called aspectual, transitive or telic SE I proposed a variationist approach to the study of SE with a specific group of verbs. The main goal of this dissertation was to answer to the following questions: What are the factors that determine the occurrence of SE with verbs of ingestion such as ‘eat’ and ‘drink’? What is the semantic/pragmatic role of SE in transitive Spanish constructions? What are the factors that determine the occurrence of the marked form over the unmarked form and vice versa when both forms are possible? What are the factors that guide the speakers’ choice of one form over the other and to what extent do the aspectual or transitivity properties of the clause trigger the use of SE in Spanish dialects?

Based on previous literature and previous pilot studies on the variable use of SE, I stated the following hypotheses: the SE that has been traditionally called an *aspectual marker or telic marker* is a *transitivity marker* that also has aspectual properties. I argue that a transitivity model such as Hopper and Thompson’s (1980) can serve as the theoretical framework to prove or reject this hypothesis.

In summary, results proved the Hypothesis that SE is a Transitivity marker and the factors that determine the occurrence of SE with verbs of ingestion are grammatically and pragmatically constrained, as shown in Table 21. This table also shows that Aspectual SE is better analyzed as a Transitivity marker; the presence of SE is determined by subject, object and verb-clause associated factors that cannot be captured by the aspectual model but can be explained through the Transitivity Model proposed by Hopper and Thompson (1980).

Factors		Favors	Disfavors
<i>Non-linguistic factors</i>			
Register		Oral	Written
<i>Linguistic factors</i>			
Verb-clause associated factors	Verb type	Comer, tomar	beber
	Mood	Imperative, conditional, subjunctive	Indicative*
	Aspect	Progressive, perfectives	Imperfectives, other
	Polarity	Affirmative*	Negative
Subject associated factors	Animacy	Non-human common nouns, 3 <sup>rd</sup> per. pron., proper names	1 <sup>st</sup> and 2 <sup>nd</sup> per. pron., human common nouns
	Grammatical number	Singular	Plural
Object associated factors	Edibility/drinkability	Non-edible/drinkable objects	Edible/drinkable objects
	Specificity	Specific	Non-specific
	Definiteness	Definite	Indefinite
	Object position	Pre-verbal	Post-verbal
Factor groups not selected as significant: dialect, grammatical person (subject), clause type, grammatical number (object)			

**Table 21. Summary of results. Linguistic and non-linguistic factors that favor and disfavor SE marking**

These results show that there are factors such as counter-expectations that constrain variable SE marking and that are not considered in the Transitivity Model, but that can be incorporated to the analysis through a quantitative analysis.

### **5.1 Analysis of SE as a transitivity marker: An alternative for grammar**

In the previous chapter I showed how the factors that constrained SE variation could be explained through a transitivity model such as that proposed by Hopper and Thompson. As was predicted in the hypotheses, SE-marked constructions were associated with the transitivity parameters that correspond to high transitivity.

The results show that, with the exception of one parameter, the presence of SE is associated with high transitivity indicators, which supports the argument that SE is linked to highly transitive clauses and it tends to occur when the high transitivity indicators occur.

For the parameter *participants*, no coding was necessary, since variable SE obligatorily required two arguments- subject and object- so, by default, all clauses were considered high in transitivity for this parameter. Following Hopper and Thompson's definition of the parameter B, Kinesis: "Actions can be transferred from one participant to another; states cannot. Thus something happens to Sally in I hugged Sally but not in I like Sally" (Hopper and Thompson 1980: 252), it was also considered unnecessary to code for this parameter since all transitive uses of consumption events refer to actions as

opposed states. Table 15 presents a summary of the findings associated with the TH and AH.

*Aspect* and *punctuality* are associated with telic and punctual predicates in Hopper and Thompson's model. My results clearly revealed that SE disfavors imperfectives and favors everything that is not imperfective, a result that supports the Transitivity Hypothesis (TH) and the Aspectual Hypothesis (AH). A close analysis of the factors that favor SE reveals that preterits favor the use of SE- a result that provide evidence for the Aspectual Hypothesis. *Aspect* came out as a significant factor only for the overall analysis and for *comer*, the reason why *aspect* did not came out as significant for 'drink' *tomar/beber* has to do with specific properties of the verb:

"eat' rather than 'drink' involves a change in the state of the food being eaten, from a whole to small crushed pieces. As such, the 'eat' verb is strongly perfective in nature. 'Drink on the other hand does not involve the same kind effect on the liquid being drunk.'" (Newman 2009: 7).

*Volitionality* and *Agentivity* were operationalized through grammatical number, assuming that first person was going to be more volitional than second and third person, and the same was done for singular and plural subjects. Although the results did not strictly follow the predictions (first person favored in SE constructions), results partially

support these parameters because singular subjects favor SE and plural subjects disfavor it. Assuming that singular objects are more individuated and, as a result, more volitional and potentially more agentive (Agent high in potency, c.f Hopper and Thompson), I claim that SE is preferred with volitional and highly agentive subjects.

Results were straightforward for the parameter affirmation; affirmative clauses favor SE and negative clauses disfavor it, as predicted by the TH.

Grammatically, SE indicates that there is a deviation from canonical transitivity in the clause. This deviation is associated with the parameters that correspond to high transitivity- two or more participants, telic and punctual action, affirmative clauses, object totally affected and highly individuated (as indicated by highly definite and specific objects).

Results for all the parameters associated with object properties clearly support the transitivity idea that SE will more likely occur with highly definite and individuated objects as predicted for the Affectedness of object and individuation of object parameters.

These results provide enough evidence to suggest that SE is associated with highly transitive clauses, as was claimed by Clements (1996) and Maldonado (2000).

To summarize I have shown that a variationist method with a transitivity approach is an accurate combination for the study of Spanish SE. The transitivity approach is the best way to explain the grammatical functions of variable SE- definiteness, mood, animacy,

aspect, polarity, grammatical subject and object position. All these factors can be clearly associated with at least one of the transitivity parameters proposed by Hopper and Thompson (1980).

<b>HIGH TRANSITIVITY</b>	<b>FAVORED BY SE</b>
A. 2 or more participants	YES
B. Kinesis	YES
C. Aspect	YES*
D. Punctuality	YES*
E. Volitionality	YES**
F. Affirmation	YES
G. Mode	NO
H. Agency	YES
I. Affectedness of O (Object)	YES
J. Individuation of O (Object)	YES
K. <i>Individuation of S (subject)</i>	YES
* Also support Aspectual Hypothesis (AH)	
** Associated with individuation of subject	

**Table 22. Analysis of results based on the Transitivity Hypothesis (TH)**

The pragmatic constraints on SE variation cannot be explained in this model, but I suggest that in terms of pragmatic constraints, SE can be explained as marker of counter-expectations of the speaker, either about the situation (in this case ingestion or consumption) or about the subject and the object involved in the event of ingestion. Table

22 illustrates how the results from this dissertation support the Hypothesis that SE is a Transitivity marker and that its presence is associated with the high transitivity features.

Through the operationalization of the Transitivity Model, it was also possible to test that SE highlights semantic properties of a specific verb type and also that these semantic properties vary across the three verbs studied. In the case of the verb type, I found that SE highlights properties such as aspect, affectedness and individuation of object and counter-expectations. Results also revealed that verbs of ingestion do have lexical specific properties that were revealed in SE marking. I found specific differences between *comer* and *beber* with respect to punctuality. These result was expected considering that ‘eat’ is perfective in nature and therefore will be more likely to occur in perfective predicates and ‘drink’ in imperfective predicates: “One difference in this respect is that ‘eat’, but not ‘drink’, provides an appropriate image for the completion of a change, just as ‘drink’ provides a source for continuity and repetitiveness” (Newman 2009: 7).

## **5.2 Non-canonical events and transitivity: An alternative for pragmatics**

These results have revealed that considering SE an aspectual marker is problematic because the completive interpretation is just one of the interpretations that can be obtained from these constructions and also because the aspectual marker approach does not account for those cases in which the sense of completion is not obtained. One of

the most revealing results presented in this dissertation is the importance of pragmatic factors, specifically what I coded as ‘edibility’. This result is relevant for many reasons. First of all, the idea of SE as a marker of counter-expectations has been discussed in the literature (Maldonado 1999), but it had only been proved quantitatively for motion intransitive verbs. These results suggest that one of the functions of aspectual, transitive or variable SE is that of marking counter-expectations of the speaker. These results confirm it; SE is more likely to occur when the speakers refer to a non-canonical event.

D’Introno, González and Rivas (2007) found the same results, although they attribute the lower rates of acceptance of (123a) to intentionality<sup>14</sup>:

- (124) a. Juan comió cuarenta pasteles él solo sin ayuda  
b. Juan se comió cuarenta pasteles él solo sin ayuda  
Juan SE ate forty cakes by himself without help

I argue that speakers prefer (123b), because it refers to a non-canonical ingestive event in which an unexpected amount of food is eaten (forty cakes) and non-canonical ingestive events are marked in Spanish with SE.

In a variationist study on variable SE marking in Spanish *salirse* ‘go out’, Torres Cacoullos and Aaron (2005) reported four contexts associated with SE marking and related to non-canonical (a-c) and habitual events (d). My results show that variable SE

---

<sup>14</sup> D’Introno, González and Rivas (2007) results show that 47% of the speakers would not say (123a) but 78% would say (123b)

marking of verbs of ingestion in Spanish occurs in contexts that are parallel to those proposed by Torres Cacoullós and Aaron (2005) (Table 23).

<b><i>SALIRSE</i></b> <b>(Aaron and Torres Cacoullós 2005)</b>	<b><i>COMERSE, TOMARSE, BEBERSE</i></b> <b>(De la Mora 2011)</b>
a. ‘exit against obstacles or rules, surreptitiously’	‘eat/drink against social rules, being forced to eat, eat something not accepted as edible’  La hacemos <u>comerse</u> el pescado ‘We force her to eat-SE the fish’  besó la tierra y comenzó a <u>comérsela</u> a puños ‘he kissed the clay and he started to <u>eat-SE</u> with his hands’
b. ‘go out abruptly’	‘eat/drink abruptly’ <u>Se bebió</u> la cerveza de un trago ‘He <u>SE drank</u> the beer in one gulp’
c. ‘leave permanently’ (a group, organization, job, home)	‘eat something up’  Lamentó <u>haberse comido</u> todos los pasteles. ‘She regretted to <u>have SE eaten</u> all the cakes’
d. ‘go out to have a good time’	‘eat to have a good time’  para irnos al bar Pau Pau a <u>tomarnos</u> una cerveza. ‘to go to the Pau Pau bar and <u>drink-NOS</u> a beer’  - decía el borracho, en voz baja -, venga a <u>tomarse</u> un traguito ‘and the drunk man said – come and <u>have-SE</u> a drink’

**Table 23. Contexts for *salir(se)* and *comer(se)*, *tomar(se)* and *beber(se)***

These results also show that highly transitive clauses and non-prototypical events are SE marked in Spanish ingestive verbs. On this basis I propose that SE has a pragmatic

function and a grammatical function, both related. In the case of the pragmatic function, SE marks that either the event or some of the participants (namely A and O) are to some extent contrary to the speaker's expectations, thus we can explain that for a person who habitually eats tongue tacos, this object is not going to be marked as such, but for a person who does not eat them it will be marked. Regarding the grammatical functions of SE, its presence reflects that the clause is highly transitive. Assuming that low transitivity clauses are prototypical transitive clauses (Thompson and Hopper 2001; Vázquez Rosas 2007), hence highly transitive clauses would be considered non-prototypical or in more general terms non-canonical. All these meanings can be overtly marked, as in the case of mood and object position (imperative, and subjunctive are non-prototypical moods, preverbal objects) or be part of the referential properties of the participants of the event or the event itself.

Although this idea requires further explanation and investigation, it is possible that the factors selected as significant as well as their hierarchy are determined by the verb type. This claim is made based on the evidence that the factors that favored SE marking were almost the same for all verbs, but they show different factor arrangements. For example in the case of *comer*, definiteness is ranked higher than edibility, thus we can expect that if there is a non-edible but indefinite object, it is possible that it will occur

without the SE in order to avoid a grammatical violation for definiteness<sup>15</sup> as we see in (125a):

- (125) a. Yo tmb (sic) comí *croquetas de perro* para saber a que sabían  
‘I also ate *dog treats* to see how they tasted’  
([www.facebook.com](http://www.facebook.com))
- b. oh si, me comí *una croqueta de perro* (marca pedigree) y te dire que saben muy mal  
‘Oh yeah, I also ME ate a dog treat and I can say that they taste bad’  
([mx.answers.yahoo.com](http://mx.answers.yahoo.com))

In the case of (126) and (127) the grammatical restriction is not violated, since the object is definite, therefore we can expect SE to occur:

- (126) el día de hoy comí *tacos de sesos* y queso de Zacatecas  
‘today I ate *brain tacos* and Zacatecas cheese’  
([www.facebook.com](http://www.facebook.com))
- (127) Hoy me comí *unos tacos de sesos*  
‘today I ME ate some brain tacos’  
([blog.wordpress.com](http://blog.wordpress.com))

As I mentioned previously, this argument requires further investigation, because it seems possible to have a re-arrangement in the hierarchies in which pragmatics is situated

---

<sup>15</sup> All these examples were found on a web search

over grammar and unexpected events are marked even if they are indefinite and non-specific.

(128) fui y me comí *hamburguesas* que estaban re ricas pero no comi mas porque me explotaba el tiroo alto y vos me ...  
'and then I ME ate hamburgers, and they were delicious but I did not eat more because I was already full'

(www.fotolog.com/es)

(129) pues como vez k me tome *tequila* ayer y me di emborache  
'as you can see I ME drank *tequila* yesterday and I got drunk'

(www.myspace.com)

(130) me tomé *ron-cola* y me marie (sic)  
'I ME drank *rum and coke* and I felt tipsy'

(www.facebook.com)

Examples such as (128) – (130) suggest that the pragmatic functions of SE (i.e. counter-expectations) are outranking the grammatical functions (i.e. definiteness), but still, more research is needed to make such a claim.

### 5.3 Counter-expectations

As was pointed out by Lyons (1994), in the last several years, there have been many studies interested in the many different ways in which speakers use language to express their perceptions, opinions and feelings, namely subjectivity, in discourse. There have been also numerous attempts to understand how the expression of subjectivity,

interacts with language structure. Explicit expression of subjectivity has been a problematic topic in linguistics. There are some elements considered to be markers of subjectivity and one of the most common of these is grammatical person, with first person singular being the prototypical site for the expression of point of view (Scheibman 2001). As Traugott suggests, subjectivity is not always overtly expressed in language “subjectivity is pervasive in language use, only a subset of elements in language serve to express it explicitly and often only in certain constructions” (Traugott 1999: 179). Following Traugott (1999) and Aaron and Torres Cacoullos (2005) I claim that counter-expectation is a matter of point of view and is therefore an example of subjectivity in language. Aaron and Torres Cacoullos (2005) show that an account of variation in the use of a grammatical morpheme can help identify the contextual factors that constitute operational measures of subjectivity (Aaron and Torres Cacoullos 2005: 629). Specifically, they argue that SE marking in Spanish *salirse* can be interpreted as a counter-expectation marker: “We can now address the nature of the subjectification process of *salirse* as a counter-expectation marker” (Aaron and Torres Cacoullos 2005: 626). They claim that the development of an ‘unexpectedness’ meaning in motion verbs-plur-REFL-marker, forms an instance of subjectification of a grammatical morpheme (Aaron and Torres Cacoullos 2005: 613).

#### 5.4 Variable SE in Spanish

The last question that I would like to discuss in this section is whether this model is able to explain variable uses of SE with other verbs. Although, more empirical studies are required. The analysis proposed in this dissertation can be extended to all variable uses of SE with all types of verbs; i.e. all cases in which the absence of SE does not alter the argument structure, grammaticality or interpretation (129-132). I claim that the pragmatic function of variable SE is that of counter-expectation marker, and the interpretation obtained from the (b) examples differs from the examples in (a) because in the (b) examples there is something unexpected about the event and the speaker is marking it with SE.

- (131) a. Nacho consiguió un trabajo envidiable  
b. Nacho se consiguió un trabajo envidiable  
'Nacho (SE) got an enviable job'

- (132) a. Adrian subió a la mesa  
b. Adrian se subió a la mesa  
'Adrian SE got up on the table'<sup>16</sup>

(Maldonado 1999:16)

I argue that in (133) the presence of SE is not necessary since the speaker is referring to a habitual event as marked by *y como todos los días* 'and as usual', therefore

---

<sup>16</sup> Translation made by me

expected. This is not the case of (134), in which SE is marking the non-habitual event- an event that occurred against the expectations of the speaker:

(133) Y como todos los días yo subí hasta el ultimo piso  
'and as usual, I SE went up to the last floor'

(134) Y un día por curiosidad, yo me subí hasta el ultimo piso  
'and one day, out of curiosity, I SE went up to the last floor'

(Torres Cacoullós and Schwenter 2008: 1456)

Although verbs such as *crecer* does not accept SE in all dialects, speakers use it when they refer to a counter-expected event. In (135) an example taken from an event posted on facebook, a speaker that is very surprised to see that many people are attending a celebration writes the following:

(135) Uy se creció la familia  
'Uy, the family SE has grown!'

(www.facebook.com)

(136) Ya se nació el chamaco  
'The baby SE was born!'

(Twitter)

I argue that these two uses of SE (135) and (136) can be also explained through the counter-expectations interpretation. In (135) the speaker uses SE to express that she is surprised to see how many people are attending to the party and in (136) the use of SE might be suggesting that the baby was born under unexpected circumstances (i.e. before she/he was expected to be born, the mother was desperate etc). The generalizations of variable SE to other verbs in which the speaker seem to be expressing opinions, feelings, perceptions etc with respect to the event support a my argument of SE being a counter-expectation marker.

Table 24 shows presents a summary of the hypotheses, predictions and results discussed in this dissertation. These results support what literature on aspectual, transitivity and telic SE claim. These results also show that current theoretical approaches are too narrow to explain the complexity of the phenomena and they also show the importance of using empirical data to reformulate current linguistic theories.

HYPOTHESIS	PREDICTION	RESULTS
Transitivity Hypothesis	SE will be favored in highly transitive clauses (associated with High Transitive Parameters) and disfavored in low transitive clauses.	SE is favored in highly transitive constructions: 2 participants, punctual events (perfectives), affirmative polarity, O totally affected and highly individuated (definite and specific).
Aspectual Hypothesis	Factors that favor the presence of SE are associated to verb-lexical properties  Perfectives will favor SE and imperfectives will disfavor SE.	Factors that determine SE marking in Spanish ingestive verbs varied depending on the lexical properties of the verb.  SE is favored in perfective predicates, although it was also favor in progressives.
Counter-expectations Hypothesis	SE will be favored in clauses in which the agent or the object, are not prototypical.	SE is favored when it occurs with non-human common nouns as a subject (non-prototypical agent) and also when it occurs with a non-edible object (non-prototypical edible object)

**Table 24. Hypotheses, predictions and results for variable SE marking in Spanish ingestive verbs**

## **5.5 Hopper and Thompson's Transitivity model revised**

The results presented in this dissertation show that the Transitivity Model can account for almost all the questions and problematic examples presented by aspectual SE but it requires some revisions in order to capture the complexity of discourse-dependent linguistic phenomena such as variable SE marking.

I have shown that the meaning of SE is not only represented at the clause level, but is also represented at a propositional level. SE has higher scope properties that the Transitivity Model cannot capture as it is formulated right now. The propositional representation of SE is related to the counter-expectations of the speaker and is pragmatically driven. In addition, there are grammatical restrictions associated with clause, verb, subject and object properties that are also constrained by the specific properties of the verb. These grammatical restrictions are responsible for the transitivity parameters associated with a specific type of verb. These lexical differences are reflected not only in the Transitivity parameters but also in the arrangement of these parameters; therefore, the parameters associated with a specific verb type (i.e. verbs of cognition) will be very similar but the arrangement of the parameters (factor groups in this dissertation) will be verb-specific. The hierarchy of factors associated with SE marking for a specific verb (in this case a verb of cognition), such as forget 'olvidarse', see 'verse' and memorize 'aprenderse' will be determined by the specific properties of the verb, while the factors selected will be determined for the 'type of verb'.

In the case of verbs of ingestion this was evident for aspect. While this factor was significant for the verb *comer* ‘eat’ it was not significant for *beber* ‘drink,’ which suggests that although SE has a close relationship with the aspectual properties of the clause it is also constrained by specific properties of the verb.

## **5.6 Conclusions**

In this dissertation I have demonstrated that what has been called aspectual SE (De Miguel and Fernandez 2000; Nishida 1994; Zagona 1996), middle SE/exploitation SE (Maldonado 1999, Maldonado 2000), telic SE (Sanz 2002) or transitivity SE (Clements, 2002), has both pragmatic and grammatical properties that cannot be fully explained by any single current theoretical approach.

Previous studies were proven to be correct on many counts, although they proved to be too narrow to explain the complexity of SE marking. SE is related to the aspectual interpretation of the clause, and it contributes to a perfective interpretation, as was suggested in previous literature (De Miguel and Fernandez de Lagunilla 2000; Nishida 1994; Sanz and Laka 2002, Zagona 1996); however, this is not the only contribution to meaning that the presence of the pronoun makes. These results also show that the presence of SE is associated with high transitivity parameters.

One of the most important contributions of this research is that it shows that Spanish SE variation is constrained by pragmatic and grammatical factors. This is true for

both transitive verbs, as was discussed in this dissertation, and for intransitive verbs, as was found in previous research.

I have proposed that variable SE (aspectual, transitive, middle, telic) is a non-canonical/counter-expectation marker that is favored in highly transitive clauses and can be explained through a revised model of transitivity that considers a propositional level that incorporates the grammatical and pragmatic contribution of SE in combination with the clause and verb associated properties (subject, object and events). In terms of grammar, SE tends to occur in non-canonical transitive clauses (associated with parameters of high transitivity) that refer to a telic and punctual event performed by volitional and highly agentive subjects that totally affect highly individuated and definite objects.

Regarding its pragmatic properties, SE is a subjectivity marker that highlights counter-expectation of the speaker with relation to the event, the subject or the object. SE marks that either the event, the subject or the object, or any combination of event, subject and object, are non-canonical with respect to what is expected (either with respect to the event or with respect to a grammatical category). In the case of ingestion events, it is expected that there will be a highly animate and volitional subject that performs the ingestion in a particular manner (i.e. sit, with specific tools) and that affects the object that is being consumed (canonically for eat and drink this object is expected to be inanimate or identified by the speaker as something 'edible or drinkable'). Thus, any deviation from these norms (eating soup with a fork, eating a human being, nails,

drinking a bottle of wine in one gulp, drinking blood etc.) will favor the presence of SE to mark the ‘non-canonical consumption event’.

The grammatical factors that constrain SE variation in Spanish can be explained through the transitivity model as proposed by Hopper and Thompson (1980) and the pragmatic factors can be explained through a model of counter-expectation, as proposed by Traugott (1999), “counter-expectation as a matter of point of view and is therefore an example of the subjectivity of language” (Traugott 1999: 179), and as was previously claimed by Aaron and Torres Cacoullos (2005) and Torres Cacoullos and Schwenter (2008) for transitive motion verbs.

This dissertation also provides evidence for SE marking and its relationship to lexical properties of the verb, which has also been tested to be the case for differential object marking (DOM) as many authors have argued (Leonetti, 2004; Von Heusinger 2008). I argue that the transitivity-revised model can also account for DOM in Spanish and it can be considered also as a counter-expectation marker. In the case of DOM, the presence of personal *a*, marks that the object is a non-canonical/counter-expected object and therefore it requires to be marked.

Although more empirical data is still required, there is enough evidence to claim that variable SE in Spanish shares the same pragmatic properties across verbs; in both motion and ingestive verbs SE tends to occur in counter-expectation contexts.

One of the advantages of the model proposed here is that it allows to explain variation of SE with verbs of ingestion including its metaphorical extensions, that for many linguists have been considered outside the field of linguistic study.

Besides the contribution to the analysis of SE, this dissertation also contributes to the study of the morphosyntactic and syntactic properties of ‘eat’ and ‘drink’, two concepts that have not received a great deal of attention in the linguistic literature to date. As has been claimed by many authors, there are important syntactic and morphosyntactic properties associated with ‘eat’ and ‘drink’ verbs across languages particularly the unusual behavior of these verbs in terms of their transitivity (Næss 2009; Amberber 2009), and as Newman states “the linguistic study of ‘eat’ and ‘drink’ concepts across languages is best served by an approach which allows for experiential, extra-linguistic realities to motivate aspects of linguistic behavior” (Newman 2009: 24).

## **5.7 Future research**

This dissertation has shown the importance and necessity of integrating variationist methods to the study of Spanish SE. One of the next steps in this research is to extend the variationist analysis to other verb classes in which SE variably occurs in order to determine if all SEs that variably occur in Spanish correspond to the same SE, or if they can be categorized in the pre-existent uses of SE, or if another interpretation should be proposed. It would also be important to include in future analyses other

ingestive verbs such as *smoke* and *swallow*. It will be also worth it to extend the variationist analysis to verbs of cognition (*olvidar(se)*, *aprender(se)*) in which variable SE also occurs and to investigate if subjectivity and transitivity can also account for the variation, and to investigate the lexical contribution of the verb in those cases.

Further future research to be derived from this dissertation is to design a way to accurately code for counter-expectation, not only for the subject or object but for the whole event and run exactly the same analysis. This will be possible through a model that incorporates counter-expectation of grammatical categories (such as subject and object) and counter-expectations of the event (encoded both grammatically and pragmatically).

Following the idea of the possible hierarchies of factors as well as rearrangement of factors grammatical>pragmatic, grammatical<pragmatic or grammatical > < pragmatic, more research is still required to determine if such a hierarchy exists and operates in both ways.

From this analysis it became evident that progressive favors the use of SE, an argument that nobody else has proposed in the literature. It will be interesting to analyze to what extent the ongoing event marked by the progressive is constraining the occurrence of SE or if some other factors, like counter-expectation are playing a role.

As it was revealed for the verb *beber* there are dialectal differences in SE marking. Unfortunately, the distribution of the data per country was not sufficient to run a more specific analysis. However, as this dissertation revealed, dialectal differences should be considered in future research.

Another contribution of this dissertation was that of showing the importance of analyzing a specific group of verbs. As in the case of motion verbs, the results presented here reveal that the lexical properties of the verb should also be considered in future analyses of Spanish SE.

One of the areas of potential future research is the study of the diachronic evolution of this particular use of SE. As it was found in a pilot diachronic study of SE marking (De la Mora 2008) it is possible that ingestion verbs in Spanish are experiencing diachronic change and it is also possible that the marked forms of verbs of ingestion become independent lexical entries restricted to transitive predicates; therefore, the non-marked form (*comer*, *beber* and *tomar*) will become the intransitive form of the verb. This phenomenon of having two different lexical entries for the same verb, one to express the transitive form and the other to express the intransitive form, has been observed in some Australian languages such as Kiribatase and Longo (c.f. Newman 2009). It is possible that this SE, which was originally a reflexive pronoun, is undergoing change, but in order to test the direction of the change, diachronic data would be needed.

It is necessary to analyze the dialectal distribution of this pronoun in depth. From my experience as a native speaker of Mexican Spanish, I have observed that there is a marked tendency on speakers to prefer SE-marked constructions over non-SE-marked constructions, especially with verbs of consumption.

Finally the variationist analysis must be extended to other verbs to determine whether the grammatical properties that constrain the variation of SE in transitive and intransitive clauses are the same or not. This research will also contribute to the understanding of the relationship between variable SE, transitivity and the lexically specific effects.

## REFERENCES

- Aaron, Jessi Eliana. 2004. The gendered use of *salirse* in Mexican Spanish: *Si me salía yo con las amigas, se enojaba*. *Language in Society*, 33, 585-607.
- Aaron, Jessi Eliana and Torres Cacoullós, Rena. 2005. Quantitative measures of subjectification: a variationist study of Spanish *salir(se)*. *Cognitive Linguistics* 16, 607-633.
- Aaron, Jessi. E. 2003. *Me salté a caminar*: Pronominal constructions with intransitive motion verbs in northern New Mexican Spanish. In Sayahi, L. (Ed.), *Selected Proceedings of the First Workshop on Spanish Sociolinguistics*, Cascadilla Proceedings Project, Somerville, Massachusetts, 123-133
- Aissen, Judith. 2003. Differential object marking: iconicity vs. economy. *Natural Language and Linguistic Theory* 21.435-483.
- Alarcos Llorach, Emilio. 1994. *Gramática de la lengua española*. Madrid: Espasa Calpe.
- Arce Arenales, Manuel. 1989. *Semantic structure and syntactic function: The case of Spanish se*. Doctoral dissertation, University of Colorado at Boulder.
- Ariel, Mira. 2008. *Grammar and pragmatics*. Cambridge: Cambridge University Press.
- Armstrong, Grant. 2009. Dissertation proposal. Georgetown University.
- Barlow, M. and Kemmer, Suzanne. 2000. *Usage-based models of language*. Stanford, CA: CSLI Publications.
- Bello, Andrés. 1847. *Gramática de la lengua castellana*, ed. by Cuervo and Alcalá-Zamora. Buenos Aires: Sopena, 1964.
- Bogard Sierra, Sergio. 2005. Aspecto, aktionsart y transitividad en español. *Nueva Revista de filología hispánica*. LIII. Num, 1. 1-29.

- Bogard Sierra, Sergio. 2006. El clítico se. Valores y evolución. In Concepción Company Company (ed.), *Sintaxis histórica de la lengua española*. La frase verbal. 755-874.
- Brown, Penelope. 2008. Verb specificity and argument realization. In *Crosslinguistic perspectives on argument structure. Implications for learnability*. Ed. By Melissa Bowerman and Penelope Brown, Lawrence Erlbaum Associates: New York. 167-189.
- Bruhn de Garavito, J, Heap, D. and Lamarche, J. 2002. French and Spanish se: Underspecified, not reflexive, *2002 CLA Proceedings*.
- Bull, William. 1965. *Spanish for teachers: Applied linguistics*. New York: Ronald
- Bybee, Joan and Rena Torres Cacoullos. 2008. Phonological and grammatical variation in exemplar models. *Studies in Hispanic and Lusophone Linguistics* 1(2), 399-413.
- Bybee, Joan, 2006. From usage to grammar: The mind's response to repetition. *Language*, 82 (4): 711-733. Cambridge University Press.
- Bybee, Joan, Revere D. Perkins and William Pagliuca. 1994. *The evolution of grammar. Tense, aspect, and modality in the languages of the world*. Amsterdam/Philadelphia: John Benjamins.
- Campos, Héctor. 1999. Transitividad e Intransitividad. In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española*. 1, Madrid: Espasa, Calpe, 1521-1574.
- Clements, J. Clancy. 2006. Transitivity and Spanish non-anaphoric se. In J. Clancy Clements & Jiyoung Yoon (eds.) *Functional approaches to Spanish syntax. Lexical semantics, discourse and transitivity*, 236-264. New York: Palgrave Macmillan.
- Company Company, Concepción. (2003). Transitivity and Grammaticalization of Object. The diachronic struggle of direct and indirect object in Spanish. In Guiliiana Fiorentino (Ed.), *Romance Objects*, New York: Mouton de Gruyter, 217-260.
- Comrie, Bernard. 1989. *Language universals and linguistic typology*. Oxford: Blackwell.
- Cristina Sánchez López (ed.), *Las construcciones con "se"*, 13-163. Madrid: Visor.

- Croft, William. 2006. *Typology and universals*. Cambridge: Cambridge University Press.
- D’Introno, Francesco, Verónica González and Javier Rivas. 2007. Aspectos sintácticos y semánticos del pronombre SE. *Boletín de Lingüística* 19(28), 5-25.
- Davies, Mark. 2002-. Corpus del español (100 millones de palabras, siglo XIII - siglo XX). Available at <http://www.corpusdelespanol.org>.
- De la Mora, Juliana. 2008. Grammaticalization of middle ‘se’ in Spanish ‘comerse’ (eat-se) constructions. From reflexive pronoun to transitivity marker, unpublished manuscript.
- De la Mora, Juliana. 2009. Variable se Marking in Spanish Ingestive Verbs: A Quantitative Analysis, paper presented at the NWA 39 Conference, Ottawa.
- De Miguel Aparicio, Elena. 1999. El aspecto léxico. In Ignacio Bosque & Violeta Demonte (dirs.) Gramática descriptiva de la lengua española, vol. 2, 2977-3060. Madrid: Espasa Calpe.
- De Miguel, Elena. and Fernandez Lagunilla, M. 2000. El Operador Aspectual SE. *Revista Española de Lingüística*, 30, 1, 13-43.
- Delancey, Scot. 1982. Aspect, transitivity and viewpoint. In P. J. Hopper (Ed.), *Tense-Aspect: Between semantics and pragmatics*, Amsterdam/Philadelphia: John Benjamins, 167-183.
- Dowty, David. 1979. *Word Meaning and Montague Grammar*, Dordrecht: Reidel.
- Dowty, David. 1991. Thematic Proto-Roles and argument selection, *Language*, 67, 3, 547-619.
- Du Bois, John W. and Sandra A. Thompson. 1991. Dimensions of a theory of information flow. Unpublished manuscript, University of California, Santa Barbara.
- Du Bois, John W. 1980. Beyond definiteness: The trace of identity in discourse. In Wallace Chafe (ed.) *The Pear Stories. Cognitive, Cultural and Linguistic Aspects of Narrative Production*, 203-274. New Jersey: Ablex Publishing Corporation.
- Finegan, Edward. 1995. Subjectivity and subjectivisation: An introduction. In Dieter Stein & Susan Wright (eds.) *Subjectivity and subjectivisation. Linguistic perspectives*, 1-15. Cambridge: Cambridge Univ. Press.

- Frajzingier, Z. and Curl T.S (Eds.). 2000. *Reflexives: forms and functions*. Amsterdam and Philadelphia: John Benjamins. 103-124.
- García, Erica. 1975. The Role of Theory in Linguistic Analysis. *The Spanish Pronoun System*. North-Holland, Amsterdam.
- Gili Gaya, Samuel. 1943. *Curso superior de sintaxis española*, Barcelona:Vox-Bibliograf.
- Givón, Talmy. 1984. *Syntax. A functional-typological introduction I*. Amsterdam/Philadelphia: John Benjamins.
- Givón, Talmy. 1987. Beyond foreground and background. In R. Tomlin (ed.) *Coherence and grounding in discourse*, 175-188. Amsterdam/Philadelphia: John Benjamins.
- Goldberg, Adele. E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Guerrero Medina, Pillar. 2005. Cardinal transitivity in foregrounding discourse: A contrastive study in English and Spanish. In Butler, Christopher, Gómez-González, María A., Doval Suárez Susana M. (eds.). *The dynamics of language use*. John Benjamins. 349-370.
- Gutiérrez Ordóñez, Salvador. 1999. Los dativos. En Ignacio Bosque y Violeta Holland.
- Heusinger, K. von, and Kaiser, G. A. 2003. Animacy, specificity, and definiteness in Spanish. In K. von Heusinger, G. A. Kaiser, & E. Stark (Eds.), *Proceedings of the workshop "Semantic and Syntactic Aspects of Specificity in Romance Languages"* (p. 41-65). Konstanz.
- Hopper, Paul. J. and Thompson, Sandra. A. (Eds.), *Frequency and the Emergence of Linguistic Structure*. New York, San Francisco, London, Academic Press.
- Hopper, Paul and Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56(2), 251-299.
- Hopper, Paul. J. 1979. Aspect and foregrounding in discourse. In Givón, T. (Ed.), *Discourse and Syntax*, Academic Press, New York, pp. 213-241.
- Hopper, Paul. J., (Ed) (1982). *Tense-aspect between semantics and pragmatics*. Amsterdam: John Benjamins.

- Hopper, Paul and Sandra A. Thompson. 1993. Language universals, discourse pragmatics and semantics. *Language Sciences* 15 (4), 357-376.
- Hopper, Paul and Elizabeth Traugott. 2003. *Grammaticalization*. Cambridge, UK: Cambridge University Press.
- Hopper, Paul. 1987. *Emergent grammar*. Berkely Linguistics Society 13, 139-157.
- Juarros Daussà, Eva. 2001. The syntactic operator "se" in Spanish. Trabajo inédito.
- Kay, P. 2004. Pragmatic aspects of grammatical constructions. In Horn, L. R., Ward, G. (Eds.), *The Handbook of Pragmatics*, Blackwell, Oxford, 675-700.
- Kemmer, Suzanne. 1993. *The middle voice*. Amsterdam and Philadelphia: John Benjamins.
- Kempchinsky, Paula. 2004. Romance se as an Aspectual Element. In J. Auger (ed). *Contemporary Approaches to Romance linguistics*, Amsterdam and Philadelphia: John Benjamins. 239-256.
- Leonetti, Manuel. 2004. Specificity and object marking: the case of Spanish a. *Catalan Journal of Linguistics*. 3: 75-114.
- Lubbers de Quesada, M. and Maldonado Soto, Ricardo. (Eds.). 2005. *Dimensiones del aspecto en español*. Universidad Autónoma de Querétaro-Universidad Nacional Autónoma de México, México.
- Lyons, J. 1994. Subjecthood and subjectivity. In *Subjecthood and Subjectivity: the status of the Subject in Linguistic Theory*, M. Yaguello (ed.) 9-17. Paris: Ophrys.
- Maldonado Soto, Ricardo. 1999. *A media voz. Problemas conceptuales del clítico se*. México: UNAM (Publicaciones del Centro de Lingüística Hispánica, 46). Morfosintaxis y Semántica. Español.
- Maldonado Soto, Ricardo. 2000. Conceptual distance and transitivity increase in Spanish reflexives. In Zygmunt Frajzyngier & Traci S. Walker-Curl (eds.) *Reflexives: Forms and functions*, 153-185. Amsterdam/Philadelphia: John Benjamins.
- Maldonado Soto, Ricardo. 2008. Spanish middle syntax: A usage-based proposal for grammar teaching. In Sabine De Knop & Teun De Rycker (eds.) *Cognitive*

- approaches to pedagogical grammar: A volume in honor of René Dirven, 155-196. Berlin/New York: Mouton de Gruyter.
- Masica, Colin P. 1976. *Defining a linguistic area: South Asia*. Chicago IL: The University of Chicago Press.
- Maslova, Elena. 2007. Reflexive encoding of reciprocity: Cross-linguistic and language-internal variation. To appear in Ekkehard König, & Volker Gast (Eds.), *Reciprocals and reflexives: Cross-linguistic and theoretical explorations*. [retrieved from <http://www.stanford.edu/~emaslova/Publications/PublicationsList.html>].
- Mendikoetxea, A. 1999. Construcciones con se: Medias, pasivas y reflejas. In Ignacio Bosque and Violeta Demonte (dirs.). *Gramática Descriptiva de la Lengua Española*. Vol. 2. Madrid: Espasa-Calpe.1635-1722.
- Miguel Aparicio, Elena de. 1999. El aspecto léxico. In Ignacio Bosque and Violeta Demonte (dirs.). *Gramática descriptiva de la lengua española*. Vol. 2, Madrid: Espasa-Calpe. 2977-3060.
- Montes Giraldo, J. J. 2003. El se del español y sus problemas. *Estudios Filológicos*, 38: 121-137 [retrieved from [http://www.scielo.cl/scielo.php?pid=S0071-17132003003800008&script=sci\\_arttext](http://www.scielo.cl/scielo.php?pid=S0071-17132003003800008&script=sci_arttext)].
- Moreno de Alba José G. 2003. Suma de minucias del lenguaje. México: Fondo de Cultura Económica. [retrieved from <http://www.fondodeculturaeconomica.com/obras/suma/r3/buscar.asp>].
- Næss, Ashlid. 2007. *Prototypical transitivity*. Amsterdam, Philadelphia: John Benjamins.
- Næss, Ashlid. 2009. How transitive are EAT and DRINK. *The linguistics of Eating and Drinking*, ed. by John Newman . Amsterdam, Philadelphia: John Benjamins. 27-43.
- Newman, John and Rice, Sally. 2006. Transitivity Schemas of English EAT and DRINK in the BNC. In G. Stefan Th. and A. Stefanowitsch (Eds.), *Corpora in cognitive linguistics: corpus-based approaches to syntax and lexis*. Berlin, Heidelberg, and New York: Mouton de Gruyter.
- Newman, John. (Ed.). 2009. *The linguistics of Eating and Drinking*. Amsterdam, Philadelphia: John Benjamins.

- Nishida, Chiyo. 1994. The Spanish reflexive clitic *se* as an aspectual class marker. *Linguistics* 32, 425-458.
- Ortiz Ciscomani, R. M. 2003. From ditransitive to monotransitive structure in the history of Spanish language. Reanalysis of objects. In M. Fortescue, E. S. Jensen, J. E. Morgensen and L. Schosler (Eds.), *Historical Linguistics 2003*, Amsterdam and Philadelphia: John Benjamins, 179-197.
- Palmer, F. R. 1986. *Mood and Modality*. Cambridge: Cambridge University Press.
- Pedersen Johan. 2004. Reflexive intensification in Spanish. Toward a complex reflexive? In M. Fortescue, E. S. Jensen, J. E. Morgensen and L. Schosler (Eds.), *Historical Linguistics 2003*, Amsterdam and Philadelphia: John Benjamins, 199-223.
- Rand, David and David Sankoff. 1990. *GoldVarb: a variable rule application for Macintosh*.
- Reuland, E. 2000. The fine structure of grammar: anaphoric relations. In Zygmunt Frajzinger and Traci S. Curl (Eds.) *Reflexives: forms and functions*. Amsterdam: Benjamins. 153-185.
- Reynoso Noverón, Jeanett. 2005. Procesos de gramaticalización por subjetivización: El uso del diminutivo en español. In David Eddington (ed.) *Selected proceedings of the 7th Hispanic linguistics symposium*, 79-86. Somerville, MA: Cascadilla Proceedings Project.
- Roldán, Mercedes. 1971. Spanish constructions with *se*. *Language Sciences*, 18. 15- 23.
- Sánchez López, Cristina. 2002. Las construcciones con *se*. Estado de la cuestión. In Cristina Sánchez López (ed.) *Las construcciones con se*, 13-163. Madrid: Visor.
- Sanz, Montserrat and Laka, Itziar.. 2002. Oraciones transitivas con *se*. El modo de acción en la sintaxis. In Sánchez López, Cristina (Ed.), *Las construcciones con se*. Visor Libros: Madrid.
- Sanz, Montserrat. 2000. Transitivity and Aktionsart. Data from Spanish. In Montserrat Sanz, *Events and predication. A new approach to syntactic processing in English and Spanish*, Amsterdam & Philadelphia: John Benjamins, pp. 39-90.
- Scheibman, Joanne. 2001. Local patterns of subjectivity. In Joan Bybee and Paul Hopper (eds.) *Frequency and the emergence of linguistic structure*, 61-89. Amsterdam/Philadelphia: John Benjamins.

- Schladt, Mathias. 2000. The typology and grammaticalization of reflexives. In Zygmunt Frajzinger and Traci S. Curl (eds.) *Reflexives: forms and functions*. Amsterdam and Philadelphia: John Benjamins. 103-124.
- Seco, Rafael. 1954. *Manual de gramática española*, ed. by Manuel Seco. Madrid: Aguilar, [1975].
- Sheibman, Joanne. 2001. Local patterns of subjectivity in person and verb type in American English conversation. In Joan Bybee and Paul Hopper (eds.), *Frequency and the emergence of linguistic structure*. Amsterdam/Philadelphia: John Benjamins. 61-89.
- Siller-Runggaldier. 2003. Changes of valence and their effect on objects. In Guiliana Fiorentino (Ed.), *Romance Objects*, New York: Mouton de Gruyter, 188-260.
- Smith, Carlota S. 1997. *The parameter of aspect*. Dordrecht: Kluwer Academic Publishers.
- Strauss, Susan. 2003. Completive aspect, emotion, and the dynamic eventive: The case of Korean V-a/e pelita, Japanese V-te shimau, and Spanish se. *Linguistics* 41(4), 653-679.
- Tagliamonte Sally. 2006. *Analyzing sociolinguistic variation*. New York: Cambridge University Press.
- Thompson, Sandra and Hopper, P.J. 2001. Transitivity, clause structure and argument structure: Evidence from conversation. In J. Bybee and P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure*, Amsterdam/Philadelphia: John Benjamins, 27-60.
- Thompson, Sandra A. 2002. Object complements and conversation: Towards a realistic account. *Studies in language* 26(1), 125-163. [Reed. as Hacia una explicación de los “complementos objeto” y la conversación. In Mercedes Sedano, Adriana Bolívar & Martha Shiro (eds.) *Haciendo lingüística. Homenaje a Paola Bentivoglio*, 347-392. Caracas: Universidad Central de Venezuela, 2006].
- Tippets, Ian and Scott A. Schwenter. 2007. Relative animacy and differential object marking in Spanish. Presented at 36 New Ways of Analyzing Variation. Philadelphia.

- Torres Cacoulios, Rena and Scott Schwenter. 2008. The structure of variable middle marking in Spanish subir(se) 'go-up' and bajar(se) 'go down.' *Journal of Pragmatics*, 40, 1455-1477.
- Traugott, Elizabeth Closs. 1999. The rhetoric of counter-expectation in semantic change: A study in subjectification. In Andreas Blank, & Peter Koch (eds.) *Historical semantics and cognition*, 177-196. Berlin/New York: Mouton de Gruyter.
- Traugott, Elizabeth Closs. 2008. The status of onset contexts in analysis of micro-changes, manuscript.
- Traugott, Elizabeth Closs and Richard B. Dasher. 2002. *Regularity and semantic change*. Cambridge: Cambridge University Press.
- Turley, J. S. 1998. A prototype analysis of Spanish indeterminate reflexive constructions. *Language Sciences* 20, 137-162.
- Vann, R. E. 1993. Middle Voice, 'No fault se' and the antiacusative: Ergativity in Spanish. In Michael Lee Mazzola (ed.), *Issues and Theory in Romance Linguistics: Selected Papers from the Linguistic Symposium on Romance Languages*, Washington, D.C. : Georgetown University Press.
- Vázquez Rozas, Victoria. 2007. A usage-based approach to prototypical transitivity. . In Nicole Delbeque and Bert Cornillie (eds.). *Interpreting constructing schemas. From action and motion to Transitivity and Causality*. Mouton De Gruyter, NY. 17-37
- Vendler, Zeno. 1957. Verbs and times. *The Philosophical Review* LXVI, 143-160. Reprinted in Z. Vendler, *Linguistics in philosophy*, 97-121. New York: Cornell University Press, 1967.
- Verhaar, J. 2000. How transitive is intransitive? *Studies in Language*, vol. 14(1), 93-168.
- von Heusinger, Klaus and Georg Kaiser. 2003. The Interaction of Animacy, Definiteness and Specificity in Spanish. In *Proceedings of the workshop: Semantic and Syntactic Aspects of Specificity in Romance Languages*. Klaus von Heusinger and Georg Kaiser eds. Konstanz: Universität Konstanz. 41-65.
- von Heusinger, Klaus and Georg Kaiser. 2005. The evolution of differential object marking in Spanish. In *Proceedings from the Workshop Specificity and the Evolution/ Emergence of Nominal Determination in Romance*. Klaus von Heusinger,

- Georg A. Kaiser and Elisabeth Stark eds. Konstanz: Universität Konstanz. 33-69.
- Wallace, Stephen. 1982. Figure and ground: The interrelationships of linguistic categories. In P. J. Hopper (Ed.), *Tense-Aspect: Between semantics & pragmatics*, Amsterdam/Philadelphia: John Benjamins, 201-223.
- Winters, R. 2006. Transitivity and the syntax of inalienable possession in Spanish. In J. C. Clements & J. Yoon (Eds.), *Functional Approaches to Spanish Syntax*, Palgrave-Macmillan: Hampshire and New York, 151-160.
- Yamamoto, Mutsumi. 1999. *Animacy and Reference*. Amsterdam/Philadelphia: John Benjamins.
- Zagona, Karen. 1996. Compositionality of aspect: Evidence from Spanish aspectual se. In C. Parodi et. al. (eds.) *Aspects of Romance Linguistics: Selected Papers from the Linguistic Symposium on Romance Languages XXIV*. Washington, DC: Georgetown University Press.