Tracing the development of Spanish participial constructions: An empirical study of semantic change

Cristina Sánchez Marco

TESI DOCTORAL UPF / ANY 2012

DIRECTORS DE LA TESI
Dr. Josep Maria Fontana (Universitat Pompeu Fabra)
Dr. Stefan Evert (Technische Universität Darmstadt)
Dr. José Luis Mendivil (Universidad de Zaragoza)
Departament Traducció i Ciencies del Lenguatge

For José

Time has fallen asleep in the
afternoon sunshine.

Ray Bradbury, *Fahrenheit 451*

Not everything that counts can be
counted, and not everything that
can be counted counts.

Albert Einstein

Acknowledgements

First of all, I would like to thank my advisors. José Luis Mendívil encouraged me to pursue this project since the beginning. His support and understanding towards my research interests has been essential to me during these years. Many thanks to Stefan Evert, who has always made me understand statistics and quantitative research as a fun and easy way to study language. I would like to thank my advisor at Universitat Pompeu Fabra, Josep Maria Fontana. Josep Maria, apart from explaining to me the fundamentals of historical linguistics, has taught me maybe the most important things in research: how to make (research) questions and how to express ideas when I write in a clear and convincing way (although I still have to improve a lot in this respect). For this, and for many other things for which there is not enough space here, I will always be grateful.

Special thanks are also due to Berit Gehrke, who has been my advisor as well in so many ways. I have always found support and encouragement from her to follow my ideas and also advice about how to formulate them better. Berit has suffered through some of the preliminary drafts of this dissertation, and has always given me insightful comments and helpful advice. Many thanks as well to Rafael Marín, who invited me to the CNRS in Lille, and with whom I have embarked on such interesting projects. My collaboration with both Berit's and Rafa's has definitely lead to a substantial improvement of this thesis and, most importantly, have made this venture a better experience.

In the corpus annotation part of this thesis I would like to thank Gemma Boleda, who helped me to tackle with all the technical challenges at the beginning of my PhD project. Thanks also to Eva Bofias, Nadjet Bouajad, Marteen Jansen, Lluís Padrò, and Oriol Valentin, who helped me in the more technical parts of this dissertation and gave me valuable comments for the corpus annotation and evaluation. Many thanks as well to Pedro Sánchez-Prieto Borja, who unselfishly provided me with the corrected versions of some of the texts in the corpus used in this dissertation.

I would also like to thank the following people for discussing various aspects of this work and/or for providing me with valuable information at different moments during this thesis: Boban Arsenijević, Núria Bel, Olga Borik, Josep Maria Brucart, Michela Cennamo, Edit Doron, José María Enguita Utrilla, Teresa Espinal, Antonio Fábregas, Ángel Gallego, Nieves Herrero, Graham Katz, Louise McNally, Jaume Mateu, Christopher Piñon, Ian Roberts, Antonella Sorace, and José Francisco Val Álvaro. Thanks as well to María Belén Villar and José Carlos de Hoyos for inviting me to talk at the Université Lyon and also for giving me honest advice, and also to the people at the Center of Cognitive Science at the University of Osnabrück, especially Peter Bosch, Carla Umbach, Stefan Hinterwimmer, and Stefan, for making me feel as if I were at my home university during

my research stay. I also thank Malte Rosemeyer and Rolf Kailuweit at Freiburg Institute for Advanced Studies and the audience and participants at the workshop *Auxiliary selection: gradience and gradualness*.

Ultimately, all of this would not have been possible without the help and support of my family and friends. My constant companion and friend at different stages of this 'intellectual' adventure, and whom I thank with great care and love, is Gabriella Lapesa. Thanks also to my friends at the Universitat Pompeu Fabra, Gemma Barberà, Stefan Bott, Mihajlo Ignjatovic, Laia Mayol, Alexandra Spalex and Berit, as well as former students Carla Parra, Héctor Martínez Alonso and Pau Giménez. I also thank to my friends Annabelle Philippe, and Roxana Sarion, for inspiring me at different moments and for letting me know that they would always be there for me. Thanks as well to Sascha Alexeyenko, Mikko Määttä, and Martin Aher who, together with Gabriella made my research stay in Osnabrück an unforgettable experience.

Thanks to my father Julio for sharing with me his creativity and for showing me how to look at the purest side of things, and my mother Rosa for showing me to be practical yet sensitive. I thank both my parents for their patiente and understanding, and their love during this time. A special thanks is also due to my uncle Alberto, who was a model to me and showed me how to think independently. I would have definitely not begun this project if it weren't for him. I thank Marta, my twin sister, and best friend: without her hospitality, generosity, care, love and artistic inspiration this thesis would have been different as well. Many thanks also to Pepe, and José, for interesting discussion about language and writing and, especially, for sharing with me their experiences and their passion for life, and to Marta, Teresa, Pilar, and Pedro as well. Lastly, I thank José for being my partner and best friend throughout this time. Our discussions during this period have always been enlightening to me. His patience and love have been unlimited.

Abstract

The main aim of this thesis is to trace the development of four different constructions involving auxiliaries and participles through the history of the Spanish language. These constructions are the perfect construction expressed by *haber* 'have' + past participle (PTCP), the verbal passive expressed by *ser* 'be' + PTCP, the adjectival passive expressed by *estar* 'be.LOC' + PTCP and the stative possessive expressed by *tener* 'tener.POSS' + PTCP.

Specifically, in this thesis I explore changes in the interpretations of these periphrases, based both on a qualitative and quantitative analysis of corpus data. I argue that these constructions have undergone a regularization change, and that this change was mainly motivated by the competition between these participial constructions for the same interpretations.

In order to test these ideas empirically, I have compiled a large diachronic corpus of Spanish from the 12th to the 20th century, consisting of more than 39 million words and composed of texts from different sources. This corpus has been automatically lemmatised and annotated with fine-grained morphosyntactic tags. In order to do this I have adapted an existing open-source linguistic analyzer (FreeLing) to allow for the efficient linguistic annotation of the oldest texts in the corpus.

Resumen

El principal objetivo de esta tesis es trazar el desarrollo de cuatro construcciones de participio, formadas por auxiliares y participios en la historia del español. Estas construcciones son la construcción de perfecto, expresada por *haber* + participio pasado (PTCP), la pasiva verbal expresada por *ser* + PTCP, la pasiva adjetival expresada por *estar* + PTCP y la estativa posesiva expresada por *tener* + PTCP.

Concretamente, en esta tesis exploro los cambios en las interpretaciones de estas perífrasis, basándome en el análisis cualitativo y cuantitativo de datos de corpus. Defiendo que estas construcciones han sufrido un cambio de regularización, y que la motivación principal para este cambio es la competición entre estas construcciones de participio para expresar las mismas interpretaciones.

Para probar estas ideas empíricamente, he compilado un corpus diacrónico del español, con textos de los siglos XII al XX, que está formado por más de 39 millones de palabras y compuesto por documentos de diferentes fuentes. Este corpus ha sido lematizado y anotado con etiquetas morfosintácticas. Para realizar esto, he utilizado una herramienta ya existente de análisis lingüístico (FreeLing), que he adaptado previamente para poder anotar de manera eficiente los documentos más antiguos que forman este corpus.

Preface

How would scribes and scholars in the Middle Ages feel about us reading and analysing all the books in their libraries in the blink of an eye? What would they think if we tell them that they can selectively retrieve and count words or expressions, for example, from chivalry books and then illustrate the behavior of such words using a graph? I guess they would probably think that it was fiction. With the increasing number of Natural Language Processing (NLP) resources and computational tools and techniques we have come closer to make this fiction story real. This dream has come true specially for English, the *lingua franca* ahead of all technological developments since the 20th century.

In this thesis I would like to contribute to this exciting and challenging task for other languages like Spanish, for which there is still much work left to do. There is an invaluable source of knowledge about history, art, medicine, science, or society in Europe written in this language in the form of manuscripts or books. Studying language change through these books is one way to contribute to the understanding, on the one hand, of this widely spoken language and, on the other, more generally, to get to know better how languages change throughout centuries. Clearly, this task is only possible to overcome by using NLP tools and statistical techniques. With the increasing number of computational tools and resources to deal with historical language varieties, such as corpora and language analyzers, properties of languages in their earliest stages can be better described and patterns of language change can be explored with greater precision. This area has undergone a quick development in the last years for languages such as English and German. With the case study presented in this dissertation I hope to contribute to this research area in Spanish as well.

Specifically, in this dissertation I focus on exploring changes in the meaning of participial constructions in the history of Spanish. The **main goal** of this thesis is to trace the emergence and development of constructions involving auxiliaries and participles analyzing the different changes these constructions underwent from a qualitative as well as a quantitative perspective. I am especially interested in understanding the semantic changes experienced by these constructions in their evolution and how the different meanings expressed by them have been differently

distributed among them across time. In particular, the empirical basis of this study are the diachronic changes of *haber* 'have' + past participle (PTCP), *ser* 'be' + PTCP, *estar* 'be.LOC' + PTCP, and *tener* 'have.POSS' + PTCP, based on a large-scale corpus of Spanish texts from the 12th to the 20th century.

How do these constructions emerge and change their interpretations over time and why? Which are the mechanisms and motivations that can explain this change? To begin to answer these questions, in this thesis I use some of the tools provided by current theoretical models in linguistics as well as different statistical methods appropriate for the analyses of these kinds of data. Thus, from the methodological point of view, I combine two approaches which are not so frequently combined in linguistics research.

Semantic change in Spanish participial constructions is an ideal case study for combining these methods, as on the one hand, recently deeper insights have been made about the meaning of these expressions, and on the other quite many hypotheses have already been proposed to explain their development. In particular, changes in the temporal and aspectual properties of periphrastic linguistic expressions in Spanish have been the focus of study specially in grammaticalization studies from the functionalist framework. However, to my knowledge this is the first study that investigates changes during the whole period of development and where hypotheses about changes in the interpretations of these periphrases are tested against quantitative data.

During my PhD and due to the lack of available resources for Old Spanish, I added a second objective to my thesis, that is the design and development of a representative diachronic corpus of Spanish and a part-of-speech tagger, which was needed in order to automatically enrich with linguistic information of lemma and morphological class the texts in this corpus.

In order to investigate language change, and before we turn to the following chapters, we need to make a basic assumption that the same principles and constraints about the semantics and composition of natural languages in synchronic languages also hold for diachronic language varieties. From this assumption it follows that similar mechanisms explaining the interpretations of linguistic expressions in modern languages should operate on historical varieties as well.

The outline of this dissertation is as follows. In chapter 1 I introduce the case study in this dissertation, that is change in the interpretations of Spanish participial constructions in Old and Modern Spanish, narrow the research questions that I will focus on answering in the following chapters, and then I present a critical review of previous approaches to this case study. After presenting the method and the composition of the corpus used to conduct this investigation in chapter 2, I move to chapter 3, where I analyse occurrences of participial constructions in the corpus and conclude with some ideas about the general type of change(s) we are observing. Chapters 4 and 5 deal with the quantitative analysis of this change.

Lastly, in chapter 6, I deal with the strategy followed to annotate the diachronic corpus used in this dissertation.I conclude this thesis in chapter 7 with a summary of the results and a suggestion for some avenues for future research.

Contents

List of Figures

List of Tables

1

THE PROBLEM

1.1 Introduction

Natural languages employ a variety of expressions to encode information about the temporal and aspectual properties of the eventualities they describe. Patterns of variation in periphrastic constructions with participles constitute an important research puzzle for linguistic theory both from a crosslinguistic and a diachronic perspective, insofar they show variation in their form and interpretation in subtle yet significant ways.

The case study of Spanish participial constructions is especially interesting in this respect. On the one hand, from the language change point of view, the emergence of periphrastic expressions is one of the most noteworthy changes that took place from Latin to the Romance languages (Bauer, 2006). Besides, from a synchronic perspective, since the earliest centuries Spanish shows a considerable number of such periphrastic expressions, which describe a wide variety of aspectual and tense distinctions and show intricate patterns of variation. For example, in Spanish there are numerous constructions to express different tense and aspectual distinctions, such as perfects, progressives, passives, and other types of secondary predications.

In this dissertation I focus on studying four participial constructions, namely *haber* 'have' + PTCP, *estar* 'be.LOC' + PTCP, *tener* 'have.POSS' + PTCP, and *ser* 'be' + PTCP, which in Modern Spanish are the expression for the perfect, adjectival passive, stative possessive and verbal passive interpretations, respectively, as will be illustrated later with examples.

This chapter has two main parts. Firstly, in Section §1.2 I present the main facts about these participial constructions both in Modern and Old Spanish which should be considered in any study about change. Secondly, in §1.3 I review previous accounts to explain change in the interpretations of participial constructions.

To conclude, I outline the contents of the successive chapters in this dissertation in Section §1.4.

Before we turn to the description of the main facts, at this point I should clarify the use of some terms. Throughout this dissertation, I will refer to abstract semantic categories such as result state, perfect or verbal passive as **interpretations**, and to the morphological instantiations of these abstract categories like *haber* + PTCP as **forms**, **constructions** or **periphrases**. I will also use the term **auxiliary** to refer to the verbs *estar*, *haber*, *ser*, and *tener* in participial constructions. Occasionally, I will refer to *ser* and *estar* in participial constructions as **copulas**. Traditionally, combinations of auxiliaries and participles have been referred to as **participial periphrases** or in Spanish **perífrasis de participio**. In these studies this term is often tied to some theoretical assumptions, such as a bi- or monoclausal and complex versus non-complex predicate analysis (Yllera, 1999). However, in this dissertation this term is used in a purely descriptive way to refer to the form of these linguistic expressions, as a synonym for **participial constructions**.

1.2 The facts

Particular semantic interpretations may or may not be licenced by a unique morphological marker or expression in a language. It is important to make a distinction between language-specific morphology and abstract categories, and study the ways in which morphology may map onto one or more abstract categories. For example, in English the simple past morphology licences both the imperfective and perfective aspectual interpretation (e.g. *painted*); whereas in Spanish there is a different morphological marker to refer to different aspectual eventualities in the past, specifically the so-called *pasado simple* or *indefinido* or simple past, e.g. *pintó* 'painted.PFV', and the *pasado imperfecto* or imperfective past, e.g. *pinta**ba*** 'painted.IPFV'. This distinction between form and interpretation will be crucial throughout this dissertation since I am interested in exploring changes in the interpretation corresponding to the same forms, that is, participial constructions, across time.

As mentioned in passing before, in Modern Spanish, similarly to the simple past example, participial constructions express abstract categories in a clear-cut way, in that each participial construction realizes one and only one interpretation. Specifically, the Spanish language morphologically distinguishes between adjectival or stative and verbal passives as well as perfects and other types of statives.

In particular, participial periphrases with *haber* instantiate the **perfect** interpretation as, informally put, they describe a past temporal interval which is somehow connected to the present, or a past situation which has present relevance (Comrie, 1976). This is illustrated in sentence (1), where the participial construc-

tion *ha vivido* describes an eventuality which extends from the past to the present, as highlighted by the use of adverb *siempre* 'always'.

(1) *Juan* ***ha*** ***vivido*** *siempre en Barcelona.*
Juan has lived always in Barcelona
"Juan has always lived in Barcelona."

On the other hand, adjectival and verbal passives are expressed by *estar* + PTCP and *ser* + PTCP, respectively. An example of an **adjectival passive** is found in (2), which describes the result state involving an entity coded as the grammatical subject. Sentence (3) is an instance of the **verbal passive** in Modern Spanish. The verbal passive interpretation of this sentence is obvious from the use of a prepositional phrase introducing an entity which bears an agentive thematic role (*por millones de ciudadanos*) and from the fact that the entity denoted by the grammatical subject is linked to a patient thematic role (*el régimen*). The fact that there is an active sentence with the same meaning as well (*Millones de ciudadanos aceptaron el régimen* 'Millions of citizens accepted the regime') further confirms this analysis.

(2) *Juan* ***está*** ***maravillado*** *de lo que sus hermanos habían sido capaces de hacer.*
Juan is.LOC amazed of the that his brothers had been able to do
"Juan is amazed at what his brothers had been capable of doing."

(3) *El régimen* ***fue*** ***aceptado*** *por millones de ciudadanos.*
the regime was accepted by millions of citizens
"The regime was accepted by millions of citizens."

In Modern Spanish there is another periphrasis involving the verb *tener*, which, similarly to *estar* + PTCP, expresses a result state.This state is often the result of some previous eventuality. However, in contrast with the periphrasis with *estar*, the entity affected by this result state is encoded as the grammatical object. *Tener* + PTCP describes a result state similarly to adjectival passives; the differences between both constructions lies basically on the fact that the former has a possessor-like subject and that the internal argument, of which the result state is predicated, is encoded as grammatical object and not as subject. This entity shows a possessive or physical tenancy or holding relation with the subject. The meaning of the constructions could be roughly paraphrased as *X has Y in state Z*, as it denotes a situation of keeping or holding an entity in a certain state. For example, in (4), *el riñón* 'the kidney'belongs to *Juan*, which is the grammatical subject of *tiene* 'has.POSS' and also the responsible entity for the kidney being spoilt.

(4) *Juan* ***tiene*** *el riñón* ***destrozado*** *de beber vino.*
Juan has.POSS the kidney spoilt of drink wine
"Juan has his kidney spoilt as he drank too much wine."

Different types of possessive relations can be expressed by this periphrasis, including inalienable possession, as in (4) and (5), or some type of construction in which the experiencer is coded as the object and the entity expressed by the grammatical subject is the presumed cause for the the eventuality expressed by the past participle , as in (6).

(5) a. *Él* ***tenía*** *el vientre* ***destrozado*** *por el mismo obús.*
He had.POSS the abdomen damaged by the same howitzer
"He had his stomach damaged by the same howitzer."

b. *El enfermo* ***tenía*** *los ojos* ***cerrados****, y respiraba*
the sick had.POSS the eyes closed and breathed
trabajosamente.
laboriously
"The sick person had his eyes closed and was breathing with great difficulty."

(6) a. *Lo* ***tienen*** *completamente* ***controlado****.*
him have completely controlled
"They have him completely under control."

b. *Esto es algo que* ***tiene*** ***abrumado****, o por lo menos*
This is something that has.POSS overwhelmed or by the least
preocupadísimo, al presidente del Gobierno.
very worried to the president of the government
"This is something by which the president of the Government is overwhelmed or, at least, very worried."

In other contexts, particularly with cognition or communication verbs, the periphrasis with *tener* denotes the result of the realisation of the event denoted by the underlying predicate with a particular emphasis or intensity, in some contexts even iterativity or repetition. In most of these cases the grammatical object may be a proposition and the possessive meaning of the construction is figurative. See examples in (7).

(7) a. ***Tengo*** ***entendido*** *que fue ella la que se marchó.*
have.POSS understood that was she the who REFL left
"I have understood that she was the one who left."

b. ***Tenemos*** ***previsto*** *invertir 40 millones en la obra.*
have.POSS planned invest 40 millions in the work

"We have planned to invest 40 millons in the work."

c. *Te* ***tengo*** ***dicho*** *que no vuelvas a casa más tarde*
you.DAT have.POSS said that not come back to home more late
de las 12 de la noche.
of the 12 of the night
"I have told you (several times) that you should not come back home later than midnight."

These uses of *tener* + PTCP have often been named under the cover names **resultative** or **stative** in the literature. However, this term is not precise enough, as it could also be used to refer to adjectival passives. Besides, in the theoretical linguistics literature **resultative** is also used to refer to a specific type of construction such as English *water the tulips flat* (Rappaport-Hovav and Levin, 2001; Kratzer, 2004; Beavers, in press, to name a few). From now on I will refer to the interpretation of Modern Spanish *tener* + PTCP as **stative possessive**, as it describes precisely both its possessive and stative nature. Occasionally, I may use the term *stative* as a cover term of the interpretations described by both *tener* + PTCP and *estar* + PTCP, as presented before.

The distribution of passives, perfects and statives, as just presented, contrasts sharply with other languages where a single morphological marker conflates different abstract categories, and other strategies have been developed to express similar interpretations. For example, in English the periphrastic participial construction with *be* realizes both verbal passives, as in (8-a), as well as adjectival passives, as in (8-b). Likewise, English expresses both perfects and stative possessives using participial constructions with *have*, see examples in (8-c) and (8-d).

(8) a. This house **was** **built** by John's father when the twins were 4.
b. That house **is** **painted** in red.
c. John **has** always **lived** in Barcelona
d. John **had** his kidney totally **destroyed** because of his binge drinking.

A different scenario for the expression of passives, statives and perfects is found in the case of German. This language is different to English and Spanish in that two different morphological markers realize the perfect; very informally stated unaccusatives form the perfect with *sein* 'be' whereas *haben* 'have' forms perfects of unergatives and transitives. See examples in (9-a) and (9-b), respectively. In contrast, German is different to English but parallels Spanish in the fact that both adjectival and verbal passives are expressed by different periphrases, specifically *sein* and *werden*, as in (9-c) and (9-d) (examples from Gehrke (2012) and Gehrke and Sánchez-Marco (2012)). As English, but differing from Modern Spanish,

stative possessives can also be expressed using the same auxiliary as in the perfect, that is *haben* (9-e).

(9) a. *Gabriella* ***ist*** *nach Osnabrück* ***gegangen***.
Gabriella is to Osnabrück gone
"Gabriella has gone to Osnabrück."

b. *Johannes* ***hat*** *einen Apfel* ***gegessen***.
johannes has an apple eaten
"Johanes has eaten/ate an apple."

c. *Die Lampe ist repariert.*
the lamp is repaired
"The lamp is repaired"

d. *Der Reifen wurde aufgepumpt, um die Fahrt fortzusetzen.*
the tire became inflated in order the journey to continue
"The tire was (being) inflated in order to continue the journey."

e. *Hans* ***hat*** *seine Niere mit Wein* ***ruiniert***.
John has his kidneys with wine destroyed
"John had his kidney destroyed because of his drinking."

Other Romance languages such as French also show patterns of variation in participial periphrases which differ from Spanish in remarkable ways. Similarly to German, French expresses the perfect using two different auxiliaries; specifically, *avoir* 'have' forms perfects of unergatives and transitives and *être* 'be' of unaccusatives, as in sentences in (10-a) and (10-b). As in English, in this language both types of passives are expressed by the same auxiliary *être* 'be', as in (10-c) and (10-d), and perfects and stative possessives are formed by the same auxiliary *avoir* (10-e).

(10) a. *Joan* ***a*** *toujours* ***vécu*** *à Barcelone.*
joan has always lived in Barcelona
"Joan has always lived in Barcelona."

b. *Joan* ***est allé*** *a Barcelona.*
joan is gone to Barcelona
"Joan has gone to Barcelona."

c. *Joan* ***est ravi*** *de Marie.*
joan is amazed of Marie
"Joan is amazed at Mary."

d. *Le régime* ***a été*** ***accepté*** *pour des millions de citoyens.*
the regime has been accepted by milliones of citizens
"The regime was accepted by millions of citizens."

e. *Joan* ***a*** ***brisé*** *les reins de boire du vin.*
joan has spoilt the kidneys of drink of wine
"Joan has his kidney spoilt as he drank too much wine."

Going back to the characteristics of participial constructions in Modern Spanish, from the morphosyntactic point of view periphrases with *haber* show some characteristics that differentiate them clearly from the other periphrases. In particular, these characteristics concern the morphological form of the participle. Whereas participles combining with *ser*, *estar* and *tener* to form passives and statives, show agreement with the internal argument, which is encoded either as grammatical subject as in (11-a) and (11-b) or as object (11-c), the participle in the Spanish perfect is always invariable. It takes always the neutral masculine singular ending *-o*, as in (11-d).

(11) a. *La reunión* ***fue considerada*** *un éxito por los*
the.F.SG meeting.F.SG was considered.F.SG a success by the
asistentes.
attendants
"The meeting was considered a success by the attendants."

b. *Los platos* ***están colocados*** *cuidadosamente en el*
the.M.PL dishes.M.PL are.LOC put.M.PL carefully in the
armario.
closet
"The dishes are carefully put in the closet."

c. *Juan* ***tiene*** *los ojos* ***enrojecidos***.
Juan has.POSS the.M.PL eyes.M.PL reddened.M.PL
"Juan has his eyes in red."

d. *Juan y yo* ***hemos empapelado*** *las paredes con un papel de*
juan and I have wallpapered the walls with a paper of
flores precioso.
flowers beautiful
"Juan and I have papered the wall with a beautiful flowered paper."

Armed with this basic description about the distribution of participial constructions in Modern Spanish, we can now take a closer look at the facts about the distribution of these constructions in the earliest centuries. In contrast with Modern Spanish, the contexts of use of participial constructions were rather different in Old Spanish. The different contextual distribution of these periphrases very clearly suggests that their interpretations were also substantially different in the earliest centuries of Spanish. As for the chronology, the received view is that approximately between the 16th and 17th centuries all these constructions were

already used similarly to Modern Spanish.

For expository reasons, it is useful to present here a concise introduction to the facts in Old Spanish and keep observations away from the analysis of such observations, which is presented in chapters 3, 4 and 5, together with more examples.

From now on, most examples used to illustrate these facts are taken from the corpus, whose contents are described in chapter 2. For each example the identification code (ID) of the text in the corpus is provided in parentheses after the translation using combinations of three or four capital letters. The complete list of titles, IDs, dates, and other information about the documents is provided in Appendix A. In cases when examples are taken from other sources, the reference will be given in the text.

Since the earliest centuries, in some contexts *haber* + PTCP holds a possessive interpretation, in a similar way to Modern Spanish participial construction with *tener*. This is illustrated in (12). The possessive meaning to the construction in the earliest centuries is not surprising, since at that time *haber* could also describe possession (Seifert, 1930), similarly to its Latin cognate *habeo*.

(12) (Stative possessive *haber* + PTCP)

a. *Ca ya el mezquino* ***auie*** *el coraçon* ***perdudo*** *& con el miedo & el pauor que auie.*
because already the mean had the hart lost and with the fear and the terror that had.
"Because the mean man had his heart lost, as he full with fear." (13th c., EE1)

b. *Las [cabras] que* ***ovieron*** *las orejas* ***tajadas*** *de pequeñas más leche an.*
the goats that had the.F.PL ears.F.PL cut.F.PL of young more milk had
"Goats which have their ears cut when they are young produce more milk." (13th c., GE1)

From the morphosyntactic point of view, *haber* + PTCP also showed some characteristics which point to an interpretation other than the perfect. In Old Spanish the participle could appear both after (as in Modern Spanish) or before *haber* at least until the 16th century, see (13-a) and (14-a), and a wide variety of elements appear between both forms, including noun phrases, as in (14-b), (14-c) and (12-a), temporal and locative modifiers (13-c), pronouns (13-a) or negation adverbials (13-b). Furthermore, in Old Spanish the participle with *haber* agrees with the object in most cases, which, as mentioned before, is a morphological characteristic of statives. This is illustrated in sentences (14-a), (14-b), (14-c), and (12-b).

(13) a. *Ya don Rachel & vidas* ***auedes*** *me* ***olbidado***
already mr Rachel and Vidas have me forgotten
"Mr Rachel and Vidas, you have already forgotten me." (12th c., CID)

b. *si podiere cobrar su muyler deue la tener assi como nuyl*
if could charge his wife must her have.POSS so as not
mal ***ouiese*** *no* ***feyto***.
evil had not done
(i) "If he could charge his wife he should have her as if she had not done any evil." (13th c., FNB)

c. *El Rey con las nouas ; ouo grant alegria .* ***Ouo***
the king with the news had great happiness had afeter moveed
luego ***mouido*** *; con la su cauallaria.*
with the his cavalry
"The king was very happy from the news. The he went away with his cavalry." (13th c., ALX)

On the other hand, in Old Spanish there were also examples such as those in (14) which are ambiguous between a stative possessive and a perfect interpretation. For example, example in (14-a) could be interpreted either a stative possessive or as a perfect. In the stative interpretation, the second person singular subject of *haber* has the entity denoted by *esta batalla* 'this battle' in a situation of being beaten. In other words, the meaning of *haber* is such that *esta batalla* could be interpreted as its object rather than as the object of *vencer* 'beat', as in the perfect interpretation. Under this interpretation the participle in the construction, *vençida* 'beaten', is a secondary predicate modifying *esta batalla*. Examples such as these, as will be shown later, are an indication that the perfect interpretation may be historically derived from a stative possessive.

(14) (Ambiguous *haber* + PTCP)

a. *Por quanto auedes fecho* ***vençida*** ***auedes*** *esta batalla*
by how much have done won.F.SG have this.F.SG battle.F.SG
(i) "You have this battle won, as you have done so much." (stative possessive)
(ii) "You have won this battle, as you have done so much." (perfect) (12th c., CID)

b. *Non* ***auia*** *el Rey;* ***acabada*** *su paraula.*
not had the king finished.F.SG his word.F.SG
(i) "The King had his speech unfinished" (stative possessive)
(ii) "The King had not finished his speech." (perfect) (13th c., ALX)

c. *Quando* ***ouo*** *la cosa;* ***dicha*** *el messagero.*
when had the.F.SG thing.F.SG said.F.SG the courier
(i) "When the courer had that thing said." (stative possessive)
(()) "When the courier had said that thing." (perfect) (13th c., ALX)

In other contexts, the possessive meaning to the construction is not so obvious, which suggests that this periphrasis could also have a perfect interpretation very close to the Modern Spanish construction. This is clear in contexts where *haber* is combined with intransitive predicates, and where there is no entity which could be interpreted as internal argument of *haber*, as in (15), or where past participles of transitives do not show agreement with the object, as in (16).

(15) (Perfect *haber* with intransitives)

a. *et se* ***auien ydo*** *a alcantara a estar enel conuento*
and REFL had gone to Alcantara to be.LOC in the convent
"And they had gone to Alcantara to stay in the convent" (13th c., GC3)

b. *QVando el rey de francia que venia detras fue entrado en*
when the king of France that came behind was entered in
Betania & ***ouo andado*** *al derredor de vn regolfo de mar...*
Betania and had walked to the surround of a gulf of sea
"When the king of France who came from behind entered Betani and had walked around a gulf in the sea..." (13th c., VLT)

(16) (Perfect *haber* with non-agreeing participle)

a. *que quando su señor houiesse de hauer la corona quelos*
that when his lord had of have the crown that the
fijos delos Reyes deuian hauer: que el que ***houiesse***
children of the kings should have that the who had
aprendido *alguna cosa por que mas valiesse.*
learnt.M.SG some.F.SG thing.F.SG by which more valued
"That when his lord ought to have the crown which the King's children should have that he had learnt something for which it would be more valuable." (13th c., BDS)

b. *les rrespondio que ellos gelo otorgaron por muchas*
them responded that they REFL it gave for many.F.PL
cosas que ***auja fecho*** *enla tierra delos moros*
things.F.PL that had done.M.SG in the land of the moorish
"They that they had given it to him because he had done many things

in the land of the Moorish." (13th c., CAX)

Interestingly, in the earliest centuries *tener* + PTCP, which began to be used productively only after the 13th century, appears in similar contexts to Old Spanish stative possessive *haber* + PTCP. Most uses of *tener* + PTCP are transitive. In these cases, the participle often agrees with the object, of which the participle describes its state and which is the possessed entity of the syntactic subject, as in (17-a) and (17-b), or the result of some communication act, as in (17-c). These cases are very similar to Modern Spanish *tener* + PTCP.

(17) (Stative possessive *tener* + PTCP)

a. *E los ojos tan luzientes & tan bermejos commo brasas & [el azor]* ***tenje*** *vnas pjuelas muy bjen* ***obradas*** *de oro*
and the eyes so shiny and so red as embers and the goshawk had.POSS some strings very well made of gold
"And the goshawk's its eyes were so shiny and red as ember and it had a string very well made in gold." (13th c., CZP)

b. *El noujo* ***tiene*** *la cabeça* ***descubierta***
the fiancé has.POSS the head uncovered
"The fiancé has his head uncovered." (13th c., CD1)

c. *El [lino] crudo* ***tiene*** *las condiciones ya* ***dichas***.
the flax raw has.POSS the conditions already said
"The raw flax has the conditions, as we already said before." (16th c., BOT)

On the other hand, in the earliest centuries it is also possible to find occurrences of *tener* + PTCP in which this periphrasis behaves more like a perfect. In these contexts the periphrasis with *tener* is combined with predicates describing some cognitive or communicative process, and whose internal argument is a proposition describing an abstract state of affairs, which can hardly be interpreted as a possessed entity. For example, in (18-a), the internal argument is a subordinate clause (*que el fijo deue semejar a su padre*); and in (18-b) and (18-c). Furthermore, in cases as in (18-b) there is an iterative adverbial (*muchas veces* 'many times') which suggests a non-stative interpretation to the periphrasis.

(18) (Perfect *tener* + PTCP)

a. ***entendido*** ***tienen*** *los omnes que el fijo deue semejar a su padre*
understood have.POSS the men that the son should seem like to his father
"Men have learnt that a son should be alike his father." (13th c.,

CD1)

b. *infeltracion dela materia ... por parte del calor y*
seeping of the matter ... for part of the heat and
corrupcion como ***tenemos*** ***visto*** *muchas vezes*
alteration as have.POSS seen many times
"The seeping of the matter because of the heat and alteration, as we have seen many times." (16th c., CBN)

c. *como arriba* ***dicho tenemos***
as above said have.POSS
"As we have said before." (16th c., BGM)

In other cases, as in (19), past participles do not show agreement with the object which, as mentioned before, is a characteristic of perfect participles in Modern Spanish. Cases such as those illustrated in (19-a), where *tener* combines with a participle which is in turn combined with a infinitive describing the beginning of a process in the past are not possible in Modern Spanish.

(19) (Perfect *tener* + PTCP)

a. *el Rey don alfonso pues que vio quela tierra del*
the king mr Alfonso because that saw that the land of the
Reyno de murçia se poblaua de xpistianos & que
Kingdom of Murcia REFL populated of christians and that
labrauan las fortalezas que ***tenjan*** ***començado a***
worked the.F.PL fortresses.F.PL that had.POSS begun.M.SG to
labrar *dexo omnes quelo fiziesen*
work left men that it did
"The King Alfonso as he saw that the land in the Kingdom of Murcia was populated by christians and that they made fortresses in which they had begun to work, he let the men do it." (13th c., CAX)

b. *et los que* ***tenian*** ***fecho*** *cueuas en los cuestos*
and the who had.POSS done.M.SG caves.F.PL in the hills
impliensse de agua
fill.REFL of water
"And the ones who had caves done in the hills were filled in with water." (13th c., GC3)

In stark contrast with Modern Spanish, where *ser* + PTCP realizes verbal passives, in Old Spanish this periphrasis could be interpreted both as a verbal and as an adjectival passive as well as a perfect with some intransitives. In this respect, it resembles, on the one hand, the distribution of modern language varieties like German, French, or Italian, which show auxiliary selection in the perfect, and on

the other, of other languages like English which do not have distinct constructions for verbal and adjectival passives.

Since the earliest centuries *ser* combines with intransitives denoting some change of state or of location such as *entrar* 'enter', holding a perfect interpretation. See examples from the 12th and 13th centuries in (20). This contrasts with Modern Spanish, when only *haber* + PTCP can be used in this context.

(20) (Perfect *ser* + PTCP)

a. ***Salidos*** ***son*** *todos armados por las torres de valençia*
gone out are all armed through the towers of Valencia
"They all have come out armed through the towers in Valencia." (12th c., CID)

b. *muchos ombres buenos de alta sangre* ***fueron*** *aqui* ***venjdos*** *para te serujr*
many men good of high blood were here come for you serve
"Many noble men had come here to serve you." (13th c., CZP)

c. *E desque* ***fue*** ***llegado*** *ala çibdat ellos non qujsieron y entrar*
and since was come to the city they not want there come
"And since he had arrived in the city, they did not want to go in there." (13th c., CAX)

Adjectival passive uses of this periphrasis in Old Spanish are clear in contexts, as in (21), in which the agent of the underlying eventuality is not mentioned and the participle is modified by degree adverbials like *muy* 'very' (21-d), which cannot appear with participles in verbal passives and perfects.

(21) (Adjectival passive *ser* + PTCP)

a. *Non se abre la puerta ca bien* ***era*** ***çerrada***
not REFL open the door because well was closed
"The door is not opened as it was well closed." (12th c., CID)

b. *Et sobreste cerco* ***son*** ***escriptos*** *los nombres de los .xij. signos .*
and about this fence are written the name of the xij signs
"And the names of the XIJ signs are written over this fence." (13th c., AST)

c. *E alexandre* ***fue*** ***marauillado*** *desto que oyo*
and Alexandre was amazed of this that heard
"And Alexandre was amazed at all which he heard." (13th c., BDS)

d. *Pensaron de folgar; ca* ***eran*** *muy* ***cansados***.
thought of rest because were very tired
"They were thinking of resting, as they were very tired."

Lastly, verbal passive uses of *ser* + PTCP are clear in contexts in which base verbs are transitives and there are event-related modifiers, such as some *por-* 'by' phrases or modifiers like *apriessa* 'quickly', which point to an agentive and non-stative interpretation. This is illustrated in sentences in (22).

(22) (Verbal passive *ser* + PTCP)

a. *Esso* ***ffue*** *apriessa* ***fecho***
that was quick done
"That was quickly done." (12th c., CID)

b. *que [los caballeros]* ***son puestos*** *por cabdiellos por mandado*
that the knights are put for leaders for command
del Rey; pora ordenar las azes de la hueste
of the King for lead the troops of the army
"And the knights are put as leaders following the King's command in order to lead the troops in the army." (13th c., ACE)

c. *la qual sentençia* ***fue*** *luego* ***Reuelada*** *por vn angel a vn frayre*
the which sentence was then revealed by an angel to a monk
agustin
augustinian
"That sentence was then revealed by an angel to an Augustinian monk." (13th c., CAS)

Similarly to *ser* + PTCP, in the earliest centuries *estar* + PTCP, which is productively used only after the 13th century (as *tener* + PTCP) could be used as a perfect with intransitives (23), as a verbal passive with some transitive predicates (24), and as an adjectival passive (25), which is the only interpretation that has survived in Modern Spanish.

(23) (Perfect *estar* + PTCP)

a. *Quando vieron los turcos aquellos vellacos que* ***estauan*** *ya*
when saw the turks those wicked that were.LOC already
llegados *al muro vinieron de todas partes*
come to the wall came of all parts
"When the turks saw those wicked men who had already come to the wall, they came from everywhere." (13th c., VLT)

b. *Et llegando a la oriella del lago non ueyendo el aun. ca a*
and arriving to the bank of the lake not seeing him yet as to

otra part ***estaua*** ***tornado***. *echo luego el manto aluenne*
other part was.LOC come back threw then the cloak far
"And upon arriving to the bank of the lake and not seeing him yet, as he had gone back to the other side, he through the cloak far away" (14th c., G2K)

(24) (Verbal passive *estar* + PTCP)

a. *Et quieren dezir algunos auctores que el consul* ***hauie estado***
and want say some authors that the consul had been.LOC
restaurado *por un sieruo ligurieno*
restored by a servant local
"And some authors want to say that the consul had been restored by a local servant." (13th c., GCI)

b. *Contra esto dize el otro que mas val* ***estar*** ***amado*** *que*
against this says the other that more worth be.LOC loved than
temjdo
feared
"Against this the other says that it is better being loved than feared." (13th c., LAT)

(25) (Adjectival passive *estar* + PTCP)

a. *E pues commo asy* ***estaran*** *aqui syenpre* ***ençerrados*** *&*
and because as so will be.LOC here always locked up and
non faran njnguna cosa dixo el cauallero çifar
not will do no thing said the knight Cifar
"And as they will be always locked up and they will do nothing, the knight Cifar said." (13th c., CZP)

b. *siella de su regno mandará escrivir pora sí este libro*
if she of her kingdom will order write for herself this book
Deuteronomio con esta ley que ***está*** ***escrita*** *en él*
Deutoronomio with this law that is.LOC written in it
"If she will order to write the Deutoronomious book of her kingdom with this law which is written in it." (13th c., GE1)

c. *Et ellos tanto* ***estauan*** ***enoiados*** *que lo touieron por bien.*
and they so were.LOC annoyed that it had for well
"And they were so annoyed that they took it well." (13th c., EE2)

This concludes the initial survey to the principal range of data that will be addressed in this dissertation. We may now consider what previous analyses have had to say about these data.

1.3 Previous analyses

A successful analysis of change in participial constructions should account for the facts just presented, which are summarised in the following points:

- Change in the interpretations of participial constructions. Following common practice in historical linguistics, I will distinguish two basic stages for change in these constructions: First, **reanalysis** or reorganization of the semantic and syntactic structure and, second, **extension**, namely the surface manifestation and spread of a succeeding reanalysis (Harris and Campbell, 1995; Langacker, 1977; Timberlake, 1977). For each participial construction this change may be summarized as:
  - The change from the stative possessive to the perfect interpretation of participial periphrases formed with *haber* and *tener*, and the extension of the former as the only expression of the perfect.
  - The generalization of use of *estar* + PTCP as the only means to express adjectival passives.
  - The loss of perfect and adjectival passive interpretations of *ser* + PTCP.
- The chronology of change, including the emergence of *estar* + PTCP and *tener* + PTCP in the 13th century texts and the fact that most changes affecting participial constructions in Spanish achieve some relevant stage between the 16th and 17th centuries, in that they come to behave similarly to the way they behave in Modern Spanish.
- The relations, motives and driving forces, that may underlie changes in the interpretations of participial constructions from Old to Modern Spanish, as it is suggested from the fact that one and only one interpretation corresponds to each participial construction in the 20th century, whereas in Old Spanish the same range of interpretations were available to more than one participial construction.

The general question addressed in previous analyses about the range of data just presented is the following: How and why do the interpretations of Spanish participial constructions change over time? In the literature different analyses have been proposed to try and answer this question about the development of participial constructions in Spanish.

Since the literature on the development of participial constructions (especially concerning the perfect *haber* + PTCP) is relatively large and has been summarized in other works, in this section I will mainly be concerned with examining and evaluating the main analytic strategies that have been adopted in the past in order

to answer this question. Furthermore, I will only review those approaches which may be directly relevant to my analysis, in that they provide quantitative evidence for their claims or they propose ideas which are related to the ones I will propose, in a way which will become clear in chapters 3, 4 and 5.

Apart from differences in the period of time that previous approaches focus on studying, they differ fundamentally on how they account for shifts in the interpretations of participial constructions and/or auxiliaries forming these periphrases. Different conditions and motivations have been argued to play an important role in these changes. Specifically, three different approaches to the development of Spanish participial constructions can be distinguished in the literature:

A. Firstly, some authors mostly from a functional-typological perspective focus on studying changes in participial constructions as instances of a quasi-universal grammaticalization process of change (Meillet, 1912; Bybee and Dahl, 1989).

B. The second approach is fundamentally distinguished by comparison with the previous one, in that it focuses on particular aspects in the development of these periphrases in Spanish, mostly in the reanalysis from the source to the target interpretation (Benveniste, 1968; Batllori and Roca, to appear), or in the extension of change (Harre, 1991; Romani, 2006; Aranovich, 2003)

C. The third group comprises those approaches that try and take a unifying perspective on the development of participial constructions in Spanish, mostly focusing on examining the relations between the uses of all or some of these constructions across time (Mendeloff, 1964; Pountain, 1985).

In this thesis, as in C, the relations underlying changes between all four participial constructions are also explored. However, the present approach differs fundamentally from C in that here I try to go one step further and investigate the particular lexical semantic factors that drive these changes and look for general patterns and mechanisms of language change underlying these changes. Furthermore, I explore change in all four participial constructions in Spanish as they occur throughout the historical time span that ranges from the 12th to the 20th century. But before I turn to present my proposal, I will present my review of previous accounts.

For the sake of clarity, it is useful to review first those approaches in the functionalist framework, which study from a typological perspective changes in participial constructions (approach A). In this part I will also introduce the main ideas about grammaticalization, which is a common label to refer to changes in the interpretation of these periphrases. Then I will turn to present those studies dealing with the development of the perfect and then with changes in passives, where both

B and C approaches are found. Some authors in B and C approaches also make use of the term grammaticalization to refer to the general change in participial constructions, however they differ from typological studies in that they focus on studying specific aspects in the development of these periphrases in Spanish.

1.3.1 Grammaticalization

It is generally assumed that a common source for perfects across languages are possessive constructions, and that the source for adjectival copulative constructions such as those expressed by *estar* + PTCP are found in verbs of posture or location. Both developments are instances of so-called grammaticalization. **Grammaticalization** refers to a general type of change from less to more functional or grammatical meanings or functions. Grammaticalization has its roots in the observation that forms with functional or grammatical meanings or uses emerge from a very basic pool or categories across languages.

A typical example of this type of change is the development of some predicates from lexical verbs to auxiliaries, as Old English *will*, which has changed from a lexical verb describing volition to be the morphological instantiation of the future tense. Other common examples are the development of serial verbs into case markers (Lord, 1973; Carlson, 1991), the change of topics into subjects (Shibatani, 1991), and the development of motion verbs into future markers as in English *be going to* (Bybee *et al.*, 1991).

Grammaticalization changes in most cases occur at the semantic, morphosyntactic and phonetic level. Morphosyntactically, in grammaticalization changes there is a change in the categorial status of the form or construction. Semantically, in grammaticalization changes the meaning of a form becomes more general; specifically, the form loses any trace of inherently referential meaning, in some cases leading to a complete loss of its lexical meaning. This has been referred to in the literature as **semantic bleaching**. In some cases, semantic and syntactic changes are also accompanied by phonetic changes, which eventually lead to a phonological reduction of the form or construction.

The categorization of the development of perfects and copulative constructions as grammaticalization phenomena goes back at least as far as Meillet (1912), and has persisted in various guises in Benveniste (1968) and Bybee and Dahl (1989), to name just a few. Participial constructions are prototypical cases of grammaticalization phenomena, in that all these constructions come to express more functional meanings or uses over time. The received view, as suggested before on the basis of the examples, is that *haber* + PTCP develops its perfect interpretation from stative possessives. On the other hand, the origins of copulas out of verbs describing a posture such as *estar* 'be.LOC' are well documented instances of grammaticalization changes.

In languages from other typological families, these changes are attested as well. For example, the development from perfect out of stative possessives takes place also in languages such as English, German, Finnish, Hindi, Bulgarian or Tamil. On the other hand, grammaticalization change from posture verbs to copulas happens also in languages nor only of the Romance family, such as Spanish, Portuguese, Catalan or French, but also in languages from other typological families, such as Kxoe, Imonda or Sango (Devitt, 1990; Heine *et al.*, 1991; Heine, 1997; Heine and Kuteva, 2002, among others).

The fact that changes of a similar kind are extensively attested in languages from different typological families has led linguists in the functionalist tradition to generalize over such changes and propose so-called **paths** or **clines** over such developments (Bybee, 1985; Bybee and Dahl, 1989; Heine *et al.*, 1991; Heine, 1997; Heine and Kuteva, 2002, among others). These crosslinguistically attested paths for change are evidence for the so-called **unidirectionality** in grammaticalization changes, which is one of the main claims of grammaticalization studies in the functionalist framework. All in all, from a semantic point of view, these clines reflect a general tendency in languages for change from from concrete to more abstract notions.

Focusing in the domain of tense and aspect systems, Bybee and Dahl (1989) argue that differences between the meanings expressed by forms across languages correspond to the location the particular form occupies along a universal path at a time. Specifically, the grammaticalization paths that may explain changes in Spanish participial constructions are the so-called perfect and copula grammaticalization clines, which I reproduce in (26) and (27), respectively (Heine and Kuteva, 2002):

(26) *Perfect grammaticalization cline*
POSSESSIVE > RESULTATIVE > PERFECT > PERFECTIVE

(27) *Copula grammaticalization cline*
STAND > COPULA

A characteristic feature of the development of the perfect in Spanish (as represented in (26)), as opposed to other languages like French or German, is that the periphrasis has not reached the last stage in the grammaticalization as a perfective marker. In some varieties of Spanish, however, there is evidence that *haber* + PTCP is developing perfective uses in some contexts. Discussion about such contexts and literature fall outside the scope of this dissertation, as here I focus on standard and European Spanish, but see Copple (2009, 2011) and Schwenter and Cacoullos (2008), among others.

As a consequence of the grammaticalization of the form, the form loses syntactic autonomy, and consequently its syntactic positioning is increasingly con-

strained; the form gradually loses its lexical content and therefore the ability to be modified independently; and also contextual restrictions about the types of subjects and verbs which can appear in the construction eventually disappear.

Based on evidence from native informants of 67 languages, Bybee and Dahl (1989) show that there are also other possible sources for the perfect and perfective across languages, namely expressions with a copula plus past participle (e.g. English, French, Finnish), or verbs meaning 'finish' (in Mandarin and Ewe), 'throw away' (as in Palaung, Korean and Fore) and movement from source (as in French, Teso, Somali or Palaung).

On the other hand, *estar* + PTCP may be analysed as a particular case of the grammaticalization of *estar* 'stand'. Change in *estar* is considered an instance of a general kind of change in which verbs describing a position, such as *stand*, *lie* or *sit* serve to develop copulative uses. Generally this change has been studied from the grammaticalization perspective as a bleaching process whereby a verb describing the position of an entity in space, for example *stare* 'stand' and *sedere* 'sit', shifts to a copula, as represented in (27).

The descriptive force of these paths or clines is clear from their typological coverage. However, when it comes to in-depth studies about specific language changes, these clines are too general and little explanatory. In particular, in Spanish they do not help to explain how or why other constructions could also give rise to the same interpretations during the same stages or why some of them eventually lose the target interpretation event if they showed at a certain stage grammatical uses which correspond to advanced positions in the clines. In other words, these clines may help to capture the source and target of change but they do not help to answer the questions of how specifically change extends over centuries or why changes in some constructions with a given interpretation survive across centuries whereas others are eventually lost.

One of the main goals of this thesis is to begin to answer these questions by investigating and exploring specific factors and constraints that underlie these very general paths of development.

Before we move to the following section, I should briefly clarify in what sense I will use the term grammaticalization in the remaining chapters. The ontological status of grammaticalization has been a matter of considerable debate during the last decade (see the special volume of *Language Sciences* edited by Campbell and Janda in 2001), specially concerning the question of whether grammaticalization can be considered a mechanism for language change on its own or whether it is the result of more general mechanisms of language change. The first view, mainly sustained by authors in the functionalist framework, is defended by authors in the studies just reviewed. The second view, mostly defended by authors in the formalist or generativist tradition, claims that grammaticalization is the result of independently motivated and more general mechanisms of language, such as reanalysis or

extension (Joseph, 2001; Newmeyer, 2001; von Fintel, 1995; Roberts and Roussou, 2003; van Gelderen, 2004; Eckardt, 2006, to name just a few). For a thorough discussion and further arguments supporting this view, see Joseph (2001).

In the analysis of change proposed in this dissertation, as presented in chapters 4 and 5, I will use the term grammaticalization as a descriptive term to refer to the result of change, which responds to more general cognitive mechanisms such as for example analogy, as will become clear in the following chapters.

Armed with the basic notions about grammaticalization, I turn now to review the literature that focus on particular aspects in the development of perfects and passives in Spanish.

1.3.2 Previous accounts of the development of the Spanish perfect

The development of the perfect, as a classical example of grammaticalization phenomena, is undoubtedly the most discussed of all Spanish participial constructions. There are two main approaches to the development of the perfect in Spanish. Some studies focus on investigating the reanalysis of stative possessives into perfects (Benveniste, 1968; Harre, 1991), whereas others focus on investigating the extension or spread of *haber* + PTCP as the perfect form from Old to Modern Spanish. The latter can be further divided in two different groups. The first group focuses on studying changes in the morphosyntactic characteristics of perfect participial constructions (Romani, 2006), and the second focuses on exploring lexical-semantic aspects which may have been relevant to the spread of the perfect with *haber*. In what follows I review these studies.

Benveniste (1968). Benveniste is one of the first authors to describe the development of perfects from Latin to Romance. He considers this development a type of conservative change, whereby a morphemic or simple expression or word is replaced by a periphrastic category with the same meaning or function, which is formed by an auxiliary and an auxiliate verb. Specifically, in Latin the perfect interpretation was instantiated by a simple or synthetic expression (e.g. *audivi* 'have heard'). This interpretation changes to be expressed by a compound expression in Romance, and eventually *audivi* restricts its value to a perfective 'heard'. Benveniste explains the development of the perfect in Romance as the result of an **auxiliation** process, whereby the combination of *habeo* 'have' and past participle comes from being a predicative construction to form a single **syntagm**.

According to this author, three conditions are required in this auxiliation process from the Latin to the Romance perfect, : First, *habeo* should be interpreted as verb of possession meaning 'have' or 'possess', not as a lexical verb meaning

'hold'. Secondly, the participle should be verbal not adjectival. Lastly, the participle must denote a sensory-intellective process inherent in the subject, such as *comperire* 'learn, discover', rather than an operational process brought on to bear on an object external to the subject.

If these three three factors converge, then the logical subject of the past participle and the grammatical subject of *habere* coincide, as they refer to the same entity. As a result of this interpretation, a new relation between the agent and the eventuality described by the participle arises, in which the agent is also interpreted as the possesor of the result state.Then, the novel temporal interpretation arises: the process is viewed as present, as the auxiliary describes, yet at the same time conceptualized as accomplished by means of the past participle. When speakers generalize this syntagmatic structure, extending it to other verbs, the paradigm of the periphrastic perfect is established. Benveniste suggests that this change would have been accomplished by the 6th century.

It will be interesting to see how in Spanish predicates describing a cognitive process may be relevant to the subsequent development of the perfect or whether other classes of predicates may explain the development of the perfect with *haber* in Spanish.

Romani (2006). In contrast with the auxiliation approach, which focus on studying the syntactic conditions that gave rise to the perfect interpretation, Romani (2006) focuses on studying the morphosyntactic changes in the construction *haber* + PTCP over time. She argues that changes in the morphosyntactic characteristics of *haber* + PTCP such as word order, participial agreement and auxiliary selection are evidence for the grammaticalization of this construction as a perfect. Romani tracks these changes in a small collection of texts from the 12th to the 15th century and shows how the construction tends to a certain stabilization of all these characteristics by the 16th century.

Tables 1.1, 1.2, 1.3 and 1.4, which I reproduce here for convenience, summarise the evidence provided by Romani in order to argue for this idea. As can be seen from these tables, word order, participial agreement, adjacency between auxiliary and participle as well as the selection of perfect auxiliary *haber*, as opposed to *ser*, gradually change towards the characteristics of the Modern Spanish perfect. Specifically, in Modern Spanish the word order of this periphrasis is invariably *haber* plus participle, very few elements can be placed between both components, the participle does not agree with the object and, lastly, the auxiliary selected to form the perfect is only *haber*. On the basis of this evidence, Romani concludes that the grammaticalization of *haber* + PTCP as a perfect seems to be completed by the 15th century, as the morphosyntactic characteristics are stabilised.

	Haber + PTCP	PTCP + *haber*
12th c.	76%	24%
13th c	97%	3%
14th c	85%	15%
15th c	99%	1%

Table 1.1: Word order of *haber* and past participle from the 12th to the 15th century (Romani, 2006).

	Haber + agreeing PTCP	*Haber* + non agreeing PTCP
12th c.	78%	22%
13th c	68%	32%
14th c	11%	89%
15th c	-	100%

Table 1.2: Participial agreement of past participles with *haber* from the 12th to the 15th century (Romani, 2006).

	Adjacent	Non adjacent
12th c.	83%	17%
13th c	88%	12%
14th c	93%	7%
15th c	93%	7%

Table 1.3: Adjacency of *haber* and past participle from the 12th to the 15th century (Romani, 2006).

	haber + PTCP	*ser* + PTCP
12th c.	70%	30%
13th c	89%	11%
14th c	78%	22%
15th c	89%	11%

Table 1.4: Frequency of *haber* + past participle vs. *ser* + past participle (as a perfect) from the 12th to the 15th century (Romani, 2006).

So far we have seen what has been said in the literature about the development of the perfect with *haber*. However, as it was presented in section 1.2, the perfect in Old Spanish, similarly to other modern language varieties like German or French, could also be expressed by *ser*. How did this copula come to be used as a perfect auxiliary, and why did this use eventually disappear in Spanish?

Traditionally, it has been assumed that the origins of the perfect uses of *ser* trace back to the paradigm of deponent verbs in Latin (see Hanssen (1912), Bastardas-Parera (1953), among others), whereas those verbs selecting the suffix *-avi* in Latin form the perfect with *haber*. In Latin deponent verbs had a periphrastic form both in the synthetic and perfect paradigms, in spite of the fact that they expressed active meanings. In contrast, periphrastic constructions formed with other verbs expressed passive meanings without exception. The idea is that intransitive predicates like Spanish *morir* 'die', *nacer* 'be born', or *caer* 'fall', formed the perfect with *ser* in Old Spanish as they are directly related to Latin deponent verbs, e.g. *morior*, *nascor* and *labor*, respectively.

However, not all verbs selecting *esse* to form perfects in Latin, that is deponent verbs, combine with *ser* in Old Spanish. There are some mismatches, mostly coming from some intransitive and reflexive verbs which seem to contradict this hypothesis. In particular, there are some verbs in Old Spanish, such as motion verbs *venir* 'come', *ir* 'go', *andar* 'walk', *exir* 'leave', which are combined with *ser* in Old Spanish despite the fact that they are not deponent verbs in Latin (*venio*, *ire*, *ando*, *exio*); or *loquor* 'speak' and *lacrimor* 'cry', which despite the fact of being deponent select *haber* in Old Spanish since the earliest centuries.

In the last years, some analyses for these facts have been proposed, mostly inspired by existing explanations on auxiliary selection in the perfect for modern varieties (Burzio, 1986; Perlmutter, 1978, 1989; Sorace, 2000). Probably on of the most influential proposals about this change in Spanish is Aranovich (2003), which I turn to review in what follows.

Aranovich (2003, 2009). Aranovich argues that the loss of the perfect uses of *ser* + PTCP as opposed to the extension of the perfect with *haber* indicates a realignment of the lexical-semantic properties that determine auxiliary selection: In Latin the selection of auxiliary *esse* in the perfect was determined by volitionality, whereas it is determined by telicity and change of location in Old Spanish. In his view, auxiliary selection in the history of Spanish is sensitive to the lexical semantics of the predicate, in particular to the affectedness of the subject. Building on Dowty's theory of proto-roles (Dowty, 1991), he claims that predicates with the more patient-like subjects resisted the spread of *haber* the longest; whereas predicates with the more agent-like subjects lost the possibility to be combined with *ser* the earliest. This author formulates this idea in the Semantic Displacement

Hypothesis, which I reproduce in (28) for convenience.

(28) *Semantic Displacement Hypothesis* (Aranovich, 2003, p. 11)
In the diachronic development of the Spanish perfect auxiliary system, the closer the subject is to being a prototypical patient, the longer the predicate resists the displacement of *ser* by *haber*.

Aranovich provides evidence for his view based on data from Benzing (1931)'s study, who observed that the replacement of *ser* by *haber* in the perfect takes place gradually and that this replacement depends on the semantic type of the participial predicate; in particular, stative verbs like *quedar* 'remain', *fincar* 'stay', and *holgar* 'rest', and existence and appearance verbs such as *acaecer* 'happen' or *aparecer* 'appear' lose their ability to combine with *ser* between the 13th and the 15th centuries. In contrast, core unaccusative such as so-called verbs of directed motion such as *ir* 'go' and *partir* 'leave' and change of state predicates like *morir* 'die' and *crecer* 'grow' are combined with *ser* at least until the 17th century.

Despite the fact that this approach seems to be reasonable to explain attested observations about how auxiliary selection with *ser* and *haber* in Old Spanish, it faces some problems that should be addressed in order to get a better understanding of change.

This approach implies that there was some kind of split intransitivity in Old Spanish, in that variation in the perfect uses of *ser* and *haber* periphrases were categorically split depending on the semantic class of the base predicate. In this respect, the Old Spanish perfect mirrors modern language varieties where there is auxiliary selection in the perfect, such as German, Italian or French. However, as will be shown on the basis of corpus data in chapter 3, the type of alternations that these periphrases show in the older stages of the language do not quite correspond to any modern language variety in this respect.

More generally, the question now may arise of how previous accounts of the development of the perfect have explained the loss of the perfect interpretations expressed by periphrases formed with *estar* and *tener* in Old Spanish. To my knowledge, there is only one study which addresses this question for *tener*, which I review in what follows. There is, however, no study, as far as I know, which explores the perfects with *estar* + PTCP in Old Spanish, thus rendering additional empirical value to this dissertation.

Harre (1991). There is general agreement in the literature that possessive *tener* follows a similar grammaticalization cline as the *haber* (Thielmann, 1885; Seifert, 1930; Pountain, 1985; Yllera, 1980, among others). The difference between both periphrases lies on the observation that *haber* + PTCP has reached an advanced stage of development as a perfect whereas *tener* + PTCP has not.

Based on Lehmann (1985)'s functional approach to grammaticalization, Harre (1991) takes the wide variety of uses of *tener* + PTCP at the synchronic level as an indication for the diachronic development in the construction. These uses include inalienable possession (29-a), resultant state uses (29-b), (29-d), and uses with communication or perception verbs (29-e). This variation found in synchronic *tener* + PTCP can be hierarchically ordered and it parallels the historical development.

(29) a. ***Tengo*** *los ojos* ***irritados***.
have.POSS the eyes reddened

b. *Juan* ***tiene*** ***preocupada*** *a su madre.*
Juan has.POSS worried to his mother

c. ***Tengo*** ***rota*** *la pierna.*
have.POSS broken the leg

d. ***Tengo*** ***pedido*** *el libro.*
have.POSS ordered the book

e. ***Tengo*** ***entendido*** *que vendrás a pasar las Navidades a casa.*
have.POSS understood that will-come to spend the Christmas at home

On the basis of a corpus study on a collection of 11 documents from the 13th to the 16th century, Harre (1991) distinguishes two different stages in the development of *tener* + PTCP. In the earliest centuries the change in *tener* + PTCP is conditioned by the lexical expansion of the uses of *tener*, which gradually covers many areas of use of *haber*. At this stage, most uses of *tener* + PTCP and *tener* express some kind of possessive relation between the participants. *Tener* gradually loses its specific semantic content of possession or of actively keeping something in a particular state, and it can in some cases be an aspectual marker of duration.

Secondly, when *tener* has already taken over the possessive uses of *haber*, *tener* + PTCP initiates its grammaticalization, where *tener* could have been analysed as an auxiliary. In these uses *tener* + PTCP gradually acquires a resultant state interpretation, in which there is some indication of previous action, and not merely a possessive state. Lastly, by the 16th century *tener* + PTCP expresses a durative or iterative action which has existed in the past and is still going at the present moment. After the 16th century, *tener* does not develop further as a perfect auxiliary, as its usage with a more restricted number of lexical items indicates.

The change of *tener* in the construction from a lexical verb meaning possession to an aspectual marker of duration explained by some semantic attrition process is essential in this grammaticalization.

Evidence provided by Harre to support her idea that there was an ongoing

grammaticalization of *tener* + PTCP in Old Spanish, up to the 16th century, are examples as in (30-a), where the instrumental phrase *con sus oncejas* 'with her nails' can only be interpreted if there is an underlying agent, therefore rendering a possible perfect interpretation to the construction. Another examples where *tener* + PTCP possibly held a perfect interpretation are in (30-b), where the periphrasis could be regarded as punctual, referring to the moment in which the anger was aroused; or in (30-c), where the stative nature of the predicate *amar* suggests that the construction could be interpreted as continuation of the past action rather than as a present state.

(30) a. ***tenié*** *con sus oncejas las massiellas* ***rompidas***
had.POSS with his nails the cheeks broken
"she had her cheeks lacerated with her nails" (Mil 364 b)

b. *La duenna pïadosa qe fue ante irada fue perdiendo la*
the lady kind that was before angered was loosing the
ira e fue más amansada; perdonólis la sanna qe lis
anger and was more calmed down forgave the brutality that her
tenié ***alzada***
had.POSS arisen
"The holy lady who had been angry before gradually grew less angry and became more calm; she pardoned them the anger which she had felt towards them" (Mil 395 a–c)

c. *sy mucho la amades mas vos* ***tyene*** ***amado***
if much her love more to you have.POSS loved
"if you love her dearly she loves you more" (LBA 798 d)

As will argued in chapters 3, 4 and 5, a better understanding of the development of the periphrasis with *tener* can be achieved if we take into account changes in the other periphrases as well. It will also be interesting to check Harre's findings about the stages in the development of *tener* + PTCP on the basis of quantitative corpus data.

1.3.3 Previous accounts of the development of the passive in Spanish

Both the extension of *estar* + PTCP as the only means to express adjectival passives and the loss of adjectival passive uses of *ser* + PTCP have been noted by previous research. However, compared to the perfect, far less attention has been devoted to the analysis of how specifically these constructions have evolved and how their interpretations changed over time.

Two main accounts have been proposed in order to account for these facts in

Spanish adjectival and verbal passives. Mendeloff (1964) focuses on describing changes in the usage frequency of *estar* + PTCP as opposed to *ser* + PTCP. The extension of *estar* + PTCP as the adjectival passive construction takes place at the expense of *ser* + PTCP, which could be interpreted as both an adjectival and a verbal passive since classical Latin. From a generativist framework Batllori and Roca (to appear), focus on investigating the reanalysis in the grammatical structure of the suceeding periphrasis *estar* + PTCP from the source to the target interpretation.

Mendeloff (1964). Mendeloff traces the evolution of the passive voice in Old Spanish from the 12th to the 15th century, based on a small-scale corpus study of the changes in the frequencies of *ser* + PTCP, *estar* + PTCP and passives expressed by verbs with a reflexive marker with *se* with different passive interpretations.

Based on the examination of changes in the usage frequencies of these constructions, he concludes that there is a change in the expression of the passive voice from the 13th to the 15th century. The factors that may have been relevant for such change affecting *ser* + PTCP are the increase in the use of the so-called passive reflexive passives with *se* during the 14th and 15th centuries, and the growth of *estar* + PTCP as means to express the adjectival passives, which reaches almost perfect concurrence with adjectival passives expressed by *ser* + PTCP by the 15th century.

On the basis of the corpus study, he concludes that the normative expression for the eventive passive in the 13th century is *ser* + PTCP, whereas in the 15th century it reverses towards the reflexive. Evidence for the first idea comes from the fact that there is enormous and very rapid increase in the use of the reflexive as opposed to similar uses of *ser* + PTCP. As can be seen from table 1.5, this increase takes place specially between the 13th and 14th centuries, when it goes from 8% (29 out of 362) to almost 50% (227 out of 460). After this century, the reflexive clearly outnumbers *ser* + PTCP in this use. In this table ratios of concurrence are derived by dividing the number of instances of the less frequent construction by the number of instances of the more frequent construction. Ratio of frequency are expressed in terms of the number of instances of concurrent constructions, reducing the lower frequency to 1. For example, as can be seen in the first row of the table, if there are 22 instances of A and 15 of B, we derive a ratio of concurrence of 0.68 (15/22) and a ratio of frequency of 1.5:1.

As for adjectival passives (or passive of state, in Mendeloff's terms), Mendeloff observes a steady increase in the frequency of use of adjectival passives expressed by *estar* + PTCP, as opposed to *ser* + PTCP, from the 12th to the 15th century. Specifically, the percentage of adjectival passives instantiated by *estar*

	A	B	Concurrence A:B	Frequency A:B
12th c.	22	15	.68	1.5:1
13th c	333	29	.09	12:1
14th c	233	227	.97	1:1
15th c	205	405	.50	1:2

Table 1.5: Total numbers and ratio of concurrence and frequency of eventive passives expressed by *ser* + PTCP (A) and reflexive passives (B) from the 12th to the 15th century (Mendeloff, 1964).

+ PTCP increases from 2 in the 12th century (1 out of 54) to yield almost 50 percentage points in the 15th century. This can be seen from Table 1.6.

	C	D	Concurrence A:B	Frequency A:B
12th c.	53	1	.02	53:1
13th c	247	33	.13	7.5:1
14th c	179	61	.34	3:1
15th c	113	106	.94	1:1

Table 1.6: Total numbers and ratio of concurrence and frequency of adjectival passives expressed by *ser* + PTCP (C) and *estar* + PTCP (D) from the 12th to the 15th century (Mendeloff, 1964).

For reasons of time, the discussion about the development of the reflexive falls out of the scope of this dissertation. However, we will see in the following chapters how *ser* + PTCP and *estar* + PTCP develop over time until the 20th century, on the basis of a large-scale corpus.

Batllori and Roca (to appear). From a generativist perpective, Batllori and Roca (to appear) argue that the extension of the adjectival passive interpretation of *estar* as well as the replacement of these uses with *ser* result from the grammaticalization of *estar* as an aspectual marker. Following Roberts and Roussou (2003)'s and van Gelderen (2004)'s generativist approach to grammaticalization, Batllori and Roca (to appear) analyse change in both Spanish and Catalan *estar* + PCTP as a shift to a grammatical element generated as a functional head. This copula changes to express the [+delimitated] feature, located in the functional projection of AspP, required to express locatives and other aspectual constructions. After this change took place, *ser* does not need to move to the higher functional projection to express this interpretation, as *estar* already supplies this function.

More specifically, they analyse change in this construction as a structural simplification (Merge instead of Move), whereby *estar* comes to merge directly in Asp with [+delimited] features, as represented in (31). In contrast, in Latin both *esse* and *stare* need to move to Asp to get those features. As opposed to *estar*, *ser* merges with a [-delimited] feature. Once *estar* was reanalysed, *ser* does not need to move to check [+delimited] features and combine with stage level predicates participles to form adjectival passives.

(31) *Syntactic representation of estar in modern Spanish*
[$_{TP}$T [$_{AspP}$ [*estar* [+del]].... [$_{VoiceP}$ Voice ..... [Small Clause ...]]]]

The motivation for this syntactic change is related to feature syncretism in the sense that the most ambiguous cue is unattended (to avoid feature syncretism of the same lexical item and ambiguity) In time *estar* becomes the more robust and and less marked option for the expression of [+delimited], and it generalizes.

The specific formulation of a syntactic treatment for these types of constructions is beyond the scope of this dissertation and for this reason I will not have anything to say about the adequacy of this analysis to capture the facts concerning the behavior of *ser* + PTCP and *estar* + PTCP at a synchronic level. However, the fact that constructions involving *ser* and *estar* can share the adjectival passive interpretation for such a long period of time and that, as we will see in chapter 4, the frequencies of appearance of *ser* + PTCP with this interpretation decrease gradually over time while the frequencies of *estar* + PTCP with that same interpretation increase gradually over time is difficult to model under an approach that views language change as an abrupt shift (or reanalysis) in the grammatical status of a given lexical expression

1.3.4 Relations between changes in Spanish perfects and passives

All accounts reviewed so far have centered around the development either of perfects or passives. We have also seen that most of them focus on studying the development of one particular participial construction at a time. If two or more participial constructions are considered, these are only periphrases formed with *haber* and *ser*, in the perfect, or with *ser* and *estar*, in passives. However, the evidence that shows that *tener* or *estar* could behave in some contexts as perfects as well, as will further corroborated in chapter 3, is too substantial to disregard. Furthermore, this fact suggest that development of both passives and perfects, which show a common source in statives, could be better understood if the development of all periphrases were studied in parallel.

However, the relations between periphrases in the development of Spanish

perfects and passives have not been the focus of interest of many studies. To my knowledge, only Pountain (1985) proposes an account for the facts about the development of perfects and passives in this line.

From a functionalist perspective, Pountain (1985) claims that there is an interdependence in changes affecting copulas and verbs of possession. Based on tabulated quantitative data from a small-scale corpus consisting of 10,000 word samples of ten texts from the 12th to the 16th century, he proposes the following hypotheses to explore such structural interdependency:

A. There is a relation between the generalization of *haber* as the only perfect auxiliary, the loss of participial agreement in the participial construction formed with this auxiliary, and the loss of the possessive interpretation to *haber* in favour of *tener*. In Pountain's words, his hypothesis is that "the more complete the dissociation of Spanish *haber* from its possessive value, the further along the path as an auxiliary it may travel". He predicts that while *haber* retains its possessive meaning, the auxiliary will combine only with transitives and the participle will agree with the object.

B. The replacement in perfect and adjectival functions of Old Spanish *ser* + PTCP, by periphrases formed with *estar* and *haber*, can be interpreted as an easing of its heavy functional load by a discrimination of these functions.

C. The generalization of *estar* + PTCP as an adjectival passive, together with the increasing use of *estar* in combination with prepositional phrases and adjectives may be visualized as related, in that they place this construction in opposition to *ser* + PTCP. He furthermore provides evidence for the establishment of adjectival passives with *estar* + PTCP in the 14th century, some time after the establishment of this copula with locative prepositional phrases but before adjectives. This is *contra* Bouzet (1953), who suggested that this did not happen until the 16th century.

In the following chapters, it will be seen how hypotheses A and C cannot be so easily corroborated on the basis of a large-scale corpus covering the whole period of development from the 12th to the 20th century. Besides, I will show how general cognitive mechanisms can account jointly for the facts covered by A, B and C more generally.

1.4 Conclusion and outline of remaining chapters

Now that we have explored the main range of data and that I have presented my review of the analyses proposed to account for these data so far, we can proceed

with my analysis. In short, in the analysis proposed in this dissertation I will try to go one step further previous approaches in two directions.

Firstly, I will base my conclusions on an analysis of all occurrences of participial constructions in a large-scale corpus covering the whole period of development in Spanish (from the 12th until the 20th century).

Secondly, I will propose an account for the motivation behind the gradual change in the rates these constructions are used with the given meanings across time. This account is based in general cognitive mechanisms such as analogy or priming and the idea that change in these constructions spreads through the lexicon, first when auxiliaries combine with some semantic classes of verbs and later with others.

But before presenting this analysis, in chapters 3, 4 and 5, in the following chapter I will present the methodology and the diachronic corpus of Spanish, which has been the empirical basis for this dissertation.

2

METHODOLOGY

2.1 Introduction

This thesis is an empirical study of semantic change. In this thesis I combine methods and tecniques from both qualitative and quantitative research in order to study changes in the interpretation of participial constructions from Old to Modern Spanish. Empirical evidence about participial constructions is analysed in order to achieve a better understanding of the development of these periphrases. In this respect, ultimately this thesis aims to contribute to the integration of theoretical and quantitative approaches into historical linguistics.

The use of electronic corpora and other NLP resources as well as statistical techniques to analyze data is becoming an indispensable tool for linguists if they want to validate hypotheses empirically. Apart from the careful examination of data coming from introspection, the analysis of experimental as well as quantitative data are gaining favour among linguists in most theoretical frameworks. The use of these tools enables linguists to find trends that would be otherwise overlooked, as well as to validate theories on empirical evidence.

The rapid advancement of these methods in linguistics is also possible thanks to the numerous breakthroughs that there have been in the last decades in the field of NLP, including the development of statistical techniques to study language data, or different corpus representation techniques, which makes it easier and fast to access language data.

The use of corpus data is even more crucial in the field of historical linguistics, where linguists need to rely on written sources entirely, as judgments of speakers or grammatical judgments of old language varieties are not available. In historical linguistics, the increasing availability of computational resources is opening new avenues for the study of linguistic change that not that long ago would have been unthinkable. The use of quantitative data allows linguists to track specific

changes in the evolution of a language as well as to identify and describe both long- and short-term trends of change that would otherwise be very hard to trace accurately. Thus resources such as corpora and NLP tools are clearly becoming an indispensable tool enabling us to access diachronic data in an easier, faster and more efficient way than it was possible for linguists before.

Some examples of the kinds of research results made possible by incorporating currently available NLP resources and techniques to the study of the evolution of a language can be seen for instance in Han and Kroch (2000), a study of the rise of *do*-Support in English using data from the *Penn-Helsinki Parsed Corpus of Middle English*, or Sagi *et al.* (2009), which traces the semantic change of expressions such as *dog*, *deer* or *do* by comparing the density of semantic vector clusters using a corpus derived from the Helsinki corpus.

For languages other than English such as Spanish, however, the on-line resources available to the research community are rather limited. Thus, despite the quantity and quality of the documents included in electronic corpora such as *CORDE*[1] or *Corpus del Español*,[2] researchers interested in the evolution of Spanish cannot carry out the type of studies conducted on the evolution of the English language due to the fact that the diachronic corpora available for this language are scarcely annotated with linguistic information and that the range of query options is not sufficiently broad.

In order to conduct an empirical study of change in Spanish participial constructions, it was therefore a challenge for me to prepare such a corpus. Hence, as a secondary aim in this dissertation I decided to add this corpus creation component. By addressing this need, I try as well to contribute to the development of NLP resources that enable linguists to carry out empirical studies using Spanish diachronic data.

The contents of this chapter are as follows. After introducing the methodology used in this thesis in Section §2.2, I turn to present, first, the main criteria considered for corpus design in Section §2.3.1 and, second, the basic statistics about the types and number of documents composing the diachronic corpus used in this dissertation in Section §2.3.2. A detailed explanation of the method used to enrich this corpus with linguistic information as well as to represent these annotations can be found in chapter 6.

2.2 An empirical study of semantic change

Any scientific theory is closely tied to empirical findings, and it always remains subject to falsification if observations incompatible with it are found. It is a funda-

[1] `http://www.rae.es`

[2] `http://www.corpusdelespanol.org`

mental part of the scientific method that all theories must be tested against observations rather than resting solely on *a priori* reasoning or intuition. Both natural and social sciences use working hypotheses that are testable by observation and experiment. Obviously, in the field of historical linguistics, these observations cannot come from grammatical judgments or any other experimental or psychological findings, as in linguistic research on modern language varieties (Sorace and Keller, 2005). In historical linguistics more than in any other field, we need to rely on contextual clues present in particular occurrences.

The case study in this dissertation makes use of the qualitative and quantitative research methods. Qualitative methods investigate the why and how of change, focusing on the careful and in-depth analysis of small samples of data, or particular examples. Then, quantitative methods can be used to seek empirical support for the analysis proposed to explain such observations. Specifically, claims are tested against quantitative evidence coming from corpora. As will be seen in chapters 4 and 5, the process of measurement is central to the quantitative research part of this study, as it provides the fundamental connection between empirical observations and mathematical expression of quantitative relationships. The numbers will provide an unbiased and objective result that can be generalized to the larger population of the given period. In other words, from the statistical analysis of the observations we will find support for theoretical claims and we will be able to generalize to the language spoken by a particular population of a given historical period.

Hence, using quantitative methods, it is possible to give precise and testable expression to theoretical ideas. To go one step further in the use of corpus evidence to explore language change, in this dissertation I follow an approach that combines current linguistic theories, in particular formal semantics (e.g. Dowty, 1979), with quantitative methods, such as Generalized Linear Models (GLMs, see Baayen, 2008). This approach allows us to keep track of the specific conditions, constraints and motivations which are relevant for change and, at the same time, track long-term as well as short-term trends in language change.

In order to conduct this empirical study of change, I will make use of a large-scale corpus, whose composition is described in the following section. R software package is used later to explore the data.

2.3 The Corpus

2.3.1 Criteria for corpus design and representativity

Different criteria were considered in order to develop the diachronic corpus which is the empirical basis in this dissertation:

- Firstly, due to the limited availability of resources, all texts that compose corpus and the NLP tools used to prepare the corpus and obtain the data from it are open-source or easy-to-access through the Web or in existing electronic collections available at university libraries.

- For the same reason, it is also essential that source documents are available in a format which is easily available to any further processing (machine-readable) without any intermediate steps, such as OCR scanning.

- The third criterion is that all resources used and the resulting corpus should be easily reusable and sustainable by other researchers. Basically, this includes the use of annotation standards and simple representation methods. Sustainability of resources and the creation of standards has been, and still is, the focus of effort of many researchers in the last decades (Bird and Simons, 2003; Ide and Romary, 2004, 2007; Rehm *et al.*, 2009, to name a few).

- It is obvious that if one wants to carry out an empirical study of change in participial constructions considering quantitative data, all occurrences of these periphrases need to be retrieved from texts automatically. In order to do this, and due to the characteristics of texts in historical language varieties (as will be presented in chapter 6), texts in the corpus should contain linguistic information about lemma and morphological class of the word.

All in all, in order to meet these criteria, I needed, first, to collect all easily available texts, preferably already existing electronic editions of Spanish. To this point I will turn in the following section. Secondly, I needed to enrich with linguistic information all words in this collection of texts. Lastly, I also had to represent the texts in this corpus in a way that information could be easily and quickly extracted, and at the same time easily portable to other formats and platforms. The last two points are the focus of chapter 6.

Furthermore, there are other more specific criteria related to the types of texts composing the corpus. A central methodological problem which has been considered for a long time in historical linguistics is that data are often quite limited. In both syntax and semantics this problem is acute, as it is common to base claims on negative evidence based of grammaticality judgments. Thus a basic assumption is needed in order to study language change empirically: the language in a corpus represents the grammar at a certain point in time. In other words, a given corpus is a sample of the grammar of the given language, which means that if a certain expression is not contained in it, such expression is not available at the grammar of such time. From the methodological point of view it follows that if we want to reach valid generalizations and conclusions from diachronic data, we need a

representative and balanced corpus, namely it should contain a broad coverage of texts in appropriate proportions, so that the population of the period of study is represented as precisely as possible (Biber, 1993).

From the diachronic point of view, whether a corpus can be considered representative or not is a matter of debate, however I will assume here somewhat informally that a corpus is representative if it contains a considerable number of words and documents from the genres and styles which exist at a certain time in a language. A balanced corpus like the *Brown Corpus of American English* or the *British National Corpus* or *Corpus de referencia del español actual*,[3] samples texts from different genres, to have a representative view of the language at a given time. An example of a diachronic corpus containing a collection of texts that spans a multitude of genres is the *Helsinki corpus* (Kytö, 1991).

Luckily, on the one hand, due to the great amount of old documents in the Iberian Peninsula and, on the other, to the long historical and philological tradition, building a representative and balanced diachronic corpus of Spanish containing texts from different periods is a challenging, but not an impossible task. Thus, in compiling the texts for the corpus used in this thesis, I tried to preserve the variety of text-types represented for each period and include a representative proportion for each, as presented in the section that follows.

2.3.2 Composition of the corpus

The documents forming the diachronic corpus of Spanish used in this dissertation come from different sources: Data from the 12th century to the 1950s were collected from the electronic texts transcribed and compiled by the *Hispanic Seminary of Medieval Studies* (HSMS),[4] the *Gutenberg project*[5] and the *Biblioteca Cervantes*.[6] This part of the corpus has been automatically annotated with linguistic information (morphosyntactic tag and lemma), using an strategy that has consisted in expanding an already existing NLP analyzer, as described in chapter 6. Additional texts from the years 1978 to 1995 were obtained from the *Lexesp corpus*, which is lemmatised and part-of-speech tagged (Sebastián-Gallés, 2000). The full list of texts in this corpus with further information about the sourceS for each document can be found in Appendix A.

The texts composing the corpus correspond to the standard variety of Spanish spoken in the Iberian Peninsula. As it is well-known, there are many distinguish-

[3] Real Academia Española: Banco de datos (CREA) [online]. *Corpus de referencia del español actual*: `http://www.rae.es`

[4] See Corfis *et al.* (1997), Herrera and de Fauve (1997), Kasten *et al.* (1997), Nitti and Kasten (1997), O'Neill (1999), Waltman (1999), Sanchez *et al.* (2003).

[5] `www.gutenberg.org/`

[6] `www.cervantesvirtual.com/`

able elements between European and American Spanish both in the vocabulary and phonology but also morphological choices concerning tense and aspect markers. Crucially for the case study in this dissertation, in some varieties of American Spanish and Iberian Spanish spoken in Gallician the perfect interpretation is mostly expressed by means of the simple past, e.g. *comí* 'lit. ate' instead of *he comido* 'lit. have eaten'. Strikingly, there are also other areas in America and Spain where the participial construction is used in past perfective contexts. Presumably, in these varieties *haber* + PTCP has reached a more advanced stage in its grammaticalization path, as a marker of perfectivity, in a similar way to other languages like French or German. The analysis of dialectal variation between American and non-standard European Spanish falls out of the scope of this thesis. An analysis of the use and development of *haber* + PTCP in these linguistic varieties can be found in Copple (2009), Copple (2011), and Schwenter and Cacoullos (2008), to name a few references.

More specifically, this diachronic corpus of European Spanish comprises more than 43 million tokens (39 million words) from the 12th to the 20th century, distributed over 674 documents as summarised in Table 2.1. I consider **token** all instances of text contained between two white spaces, including punctuation marks. **Word** refers to all instances of text contained between two white spaces, excluding punctuation symbols. In contrast, **type** refers here to any distinct token that occurs in the corpus. In this table the first three rows indicate the distribution in percentage points of the tokens, words and types, respectively. Central columns indicate the percentages for each century. The total numbers for each category are given in the last column. As can be seen from this table, percentages range from 10% to 22% for most centuries; only the 12th, 17th and 18th centuries show a lower percentage. There are, however, bigger differences in the number of documents per century. The last row in this table indicates the number of documents per century in the corpus, which range from 1 in the 12th century to 202 in the 20th century.

Century:	12	13	14	15	16	17	18	19	20	Totals
To	0.1	21.3	11.5	17.1	11.6	2.2	0.8	16.6	18.7	43,271,721
W	0.1	21.9	12.3	17.8	11.8	2.1	0.8	15.6	17.7	39,189,962
Ty	0.4	16.4	16.0	22.8	11.2	3.2	1.9	10.0	18.2	688,653
D	1	87	83	104	89	13	4	91	202	674

Table 2.1: Percentage points and total number of tokens (To), words (W), types (Ty) and number of documents (D) in the corpus from the 12th to the 20th century.

As discussed in the previous section, building a representative corpus is not only a matter of the number of words or documents but also of the types of texts

composing the corpus. The texts forming the diachronic corpus used in this thesis have been carefully selected so that most types of texts characteristic for each period are represented. Deciding which is the genre of a given document is not a trivial task. For example, it is a well-known fact that some texts show characteristics from different genres, thus making it a challenging task to decide which is the genre that should be assigned. This task is particularly difficult for diachronic corpora, as there are obvious differences and changes in the types of texts across centuries. For example, the first press and newspapers texts go back to the 19th century and, consequently, in a diachronic corpus there could be texts of this type only from this century on. To put another example, (presumably) the first novel written in Spanish appeared by the 16th century (*El Quijote*). Before that time, there were only other types of fiction books, characterised by their linear narrative line and the use of other strategies such as nested stories or verse inside a narration. Obviously, a representative diachronic corpus of Spanish should contain both newspapers and novels; however, thus it is clear that these texts will not be homogeneously distributed during all periods. Given the well-known fact that genres do indeed influence grammatical choices, we should therefore keep track of this factor.

As it is well-known, there are basically two strategies to distinguish the genre of a given text. In the first genres are assigned externally, on the basis of the communicative purpose of texts (Swales, 1990). In the second strategy, genres are assigned internally taking into account a group of defining linguistic features for each genre (Kessler *et al.*, 1997). Recently, due to the big size of most corpora and NLP resources and increasing use of this type information in fields such as information retrieval, there has been some research to automatically detect genres in texts (Lim *et al.*, 2005; Webber, 2009; Petrenz and Webber, 2012; Remus and Bank, 2012).

In this dissertation, the first thing to consider about the types of texts included in the corpus is that all documents belong to the written register. As for the genre type, to keep things as simple as possible and for reasons of time, documents in this corpus have been classified into different genres on the basis of the genre which was previously assigned to them in the sources. When this information was not available, I took as a reference the topic classification used in *CORDE*. This was possible because most texts contained in the corpus used in this dissertation are also contained in *CORDE*.

Specifically, each text in the corpus has been classified according to its genre in one of the following ten categories: poetry, history, law, didactics, prose, religion, medicine, letters, drama, or newspapers. These categories are characterised in terms of discourse features which have been traditionally recognized as genre defining. For example, typically letters show a particular layout with addresses at the top and the use of formulaic expressions at the beginning and ending of the

contents. Prose includes all fiction texts written in prose, including novels but also other fiction prose texts such as chivalry books.

Table 2.2 summarises the genre-types of texts in percentage points per century. Bold-faced letters mark the highest percentage in each century, showing the genre that is mostly represented. It is relevant to note this here, as in the frequency study presented in chapter 4 we will see how changes in participial constructions can be explored through different types of texts. As can be seen from this table, both history and prose are the genres which are represented to a greater extent throughout most centuries, and didactics despite the fact of being not so frequent, it is however represented to some extent throughout most centuries. Percentages in these types of texts range from 30% to 50% in the case of history and from 55% to 80% in prose. In contrast, other genres such as poetry, newspapers, drama or medicine, are underrepresented over time, due to reasons similar to the ones mentioned in the newspapers and novels example.

Century:	12	13	14	15	16	17	18	19	20
Poetry	**100**	1	2.2	8.9	-	7	5.1	1.4	-
History	-	**32.9**	**49.5**	13.3	7.6	8.5	**38.5**	4.1	-
Law	-	27.7	15.2	17.8	0.3	-	-	-	-
Didactics	-	19.8	18.9	**27.9**	8.8	4.9	35.6	6.3	15.9
Prose	-	15	8.7	17.9	35.3	**60.5**	-	**84.1**	**55.9**
Religion	-	3.6	4.4	8.3	-	-	-	-	-
Medicine	-	-	1.1	5.9	**48.1**	2.8	-	-	-
Letters	-	-	-	0.1	-	-	20.7	-	-
Drama	-	-	-	-	-	16.3	-	2.5	1.1
Newspaper	-	-	-	-	-	-	-	1.5	27.1

Table 2.2: Percentage points of genres per century in the corpus.

In order to facilitate the statistical analysis of the data and to make generalizations on data easier, the corpus has also been divided into four periods, following the customary division determined by external historical events which have been used as well in the division of Ages of the History of Europe: Middle Spanish (1100-1492), Modern Spanish (1493-1788), Contemporary Spanish (1789-1974), and Late Contemporary Spanish (1975-2000). The historical milestones which serve as the basis for periodization are the following:

- 1492: The discovery of America, the invention of the printing press, the publication of Nebrija's grammar.
- 1789: The French and industrial revolution.

- 1975: The end of the dictatorship in Spain.

Table 2.3 summarises the number of documents per period and century in the corpus. To avoid confusion, when I refer to these periods I will explicitly mention that I am referring to periods, and not the general labels Old and Modern Spanish. Old Spanish, as used until now, will thus be used as a cover term for both the Middle and Modern periods; and Modern Spanish, when not mentioned explicitly, will refer to the Contemporary and Late Contemporary periods.

Century:	12	13	14	15	16	17-18	19	20
Middle	1	86	83	82	-	-	-	-
Modern	-	-	-	22	89	17	-	-
Contemporary	-	-	-	-	-	-	91	30
Late Contemporary	-	-	-	-	-	-	-	172

Table 2.3: Number of documents per period in the corpus.

2.4 Conclusion

Armed with the basic methodological assumptions and a description of the contents of the corpus, we can now turn to the core chapters in this dissertation, where I present my analysis of the development of participial constructions in Spanish. In particular, in chapter 3 I will present an analysis of the distribution of participial constructions in Old Spanish and some preliminary ideas about the type of change that may explain differences between the distribution of these periphrases in Old and Modern Spanish. Then we will turn to chapters 4 and 5, where my analysis of change is presented based on quantitative evidence.

3

THE DISTRIBUTION OF PARTICIPIAL CONSTRUCTIONS IN OLD SPANISH

3.1 Introduction

This chapter focuses on the distribution of participial constructions in Old Spanish. Prior to any proposal of the development of perfects and passives we need to survey the contexts in which participial constructions appear in Old Spanish, based on the qualitative analysis of data retrieved from a large-scale corpus, as presented in the previous chapter. Only by doing this, we will be able to establish precisely which were the interpretations of these periphrases in the earliest centuries, and to what extent periphrases formed with different auxiliaries appeared in the same contexts or not, as we presented in chapter 1. Then, we will also be able to compare these findings to the uses of these periphrases in Modern Spanish. Once the contrasts between the contexts of use of participial constructions both in Old and in Modern Spanish have been established we will begin to understand precisely which sort of change might have occurred. The change as it (gradually) happens over time from the 12th to the 20th century will be investigated in chapters 4 and 5.

The range of interpretations and their distributional characteristics considered here draws on earlier literature on perfects, passives and statives. Specifically, the contexts of use of participial constructions considered here are the following:

- **Perfect contexts** are those contexts favouring the perfect interpretation.
- **Passive contexts** are contexts of use which support the passive interpretation. These contexts come in two types: **verbal passive** and **adjectival**

passive contexts, which favour interpretations after which they are named.

- **Stative possessive contexts** bear a stative possessive interpretation, as introduced in chapter 1.

The contents of this chapter are as follows. In Sections §3.2, §3.3 and §3.4, I present, respectively, the perfect, passive and stative possessive contexts of participial constructions in Old Spanish. Then I move, in Section §3.5, to present other contexts of use of the auxiliaries *haber* 'have', *ser* 'be', and also *estar* 'be.LOC' and *tener* 'have.POSS' in Old Spanish, which will be useful to achieve a complete picture of the complexity of the Old Spanish system. After an interim summary in Section §3.6, we will turn to Section §3.7, where I will propose an analysis of the meaning of the auxiliaries in Old Spanish and some ideas about which might have been the semantic change which explains differences between Old and Modern Spanish in this area.This chapter concludes, in Section §3.8, with a general analysis of the system in the earliest centuries in comparison to the system in Modern Spanish and some ideas about which might have been the motivation for change in the interpretations of participial constructions.

3.2 Perfect contexts of participial constructions

3.2.1 Basic facts about the perfect

In order to explore the development of the perfect from Old to Modern Spanish, we need to know some basic facts about the main properties of the perfect. Informally put, the perfect is used in several languages to express a past situation with current relevance (Comrie, 1976). It denotes a past state of affairs which holds some relation with the present.

The perfect has been used as a cover term for a wide range of readings. There are three main readings to the perfect (Binnick, 1991; Comrie, 1976, 1985; Dahl, 1985; Leech, 1971; McCawley, 1971, 1981, among many others):

- **Continuative** or **universal** reading (U-reading), as in (1-a), denotes a state of affairs that holds throughout an interval from a certain point in the past to the time indicated by the tense of the auxiliary. In other words, it describes a situation that started in the past but continues or persists into the present or utterance time.

- In contrast, the **non-continuative** or **existential** (also called **experiential**) reading (E-reading), as in (1-b), indicates that a given situation has held at least once during some time in the past leading up to the present. There is at

least one event of the type denoted by the base predicate in the interval terminating at the evaluation time (Dowty, 1979; McCoard, 1978; Mittwoch, 1988; Vlach, 1993).

- Lastly, the **resultative** or **perfect of result** reading (R-reading) indicates that there is one event of the type denoted by the base sentence in the interval terminating at the evaluation point, and its direct result holds at this point. This reading has often been considered a subtype of the existential. Resultative perfects always hold the inference that the state denoted by the participle holds at speech time. An example of this perfect is in (1-c), where the sentence implies that Alexandra is in Barcelona at the present time.

(1) a. Since 2000, Alexandra **has lived** in Barcelona.
b. Alexandra **has been** in Barcelona (before).
c. Alexandra **has** (just) **arrived** in Barcelona.

However, in some contexts it is not easy to decide which is the reading of the perfect. For example, in (2) (after Pancheva (2003)) it is not clear whether the sentence holds a resultative or existential reading. On the resultative reading, it requires that the glasses be lost at the reference time, while on the E-reading there is no such requirement.

(2) I **have lost** my glasses.

As it is well-known, there are some distributional characteristics, mainly concerning the aspectual type of the base predicates and the combinations with some adverbial modifiers, which favour each of these readings and thus allow to disambiguate such readings in context (Dowty, 1979; Mittwoch, 1988; Vlach, 1993; Portner, 2003; Iatridou *et al.*, 2003; Musan, 2002; Mittwoch, 2008).

For example, universal readings are only possible with perfects built out of stative predicates, including individual-level (IL) and stage-level (SL) predicates, and progressives. It is also difficult to have a universal perfect in the absence of a temporal adverbial; relevant adverbials that enable this reading in English are *since* and *always*. See examples illustrating these points in (3):

(3) (U-reading)
a. I **have understood**.
b. Mihajlo **has been** in Barcelona since Tuesday.
c. Berit **has been swimming** since noon.

In contrast, stative as well as eventive predicates like activities, accomplishments and achievements can form existential perfects. Other context which seem

to favour this reading are indefinite temporal adverbials such as *before*, *lately*, *at X time*, and adverbs of quantity or cardinality like *twice*. See examples in (4):

(4) (E-reading)

a. John **has smoked**.
b. I **have heard** him play that piece before.
c. **Have** you ever **been** to Zaragoza?
d. I **have locked** the door twice.

Lastly, the resultative reading to the perfect is only possible with telic eventualities denoting transitions consisting of an event and a result state determined by the meaning of the base predicate. In Aktionsart terms, these predicates are accomplishments and achievements. Besides, in resultative perfects, the agent argument of the verb is often irrelevant for the result state, and there are also restrictions on manner adverbs modifying the event component of the verb's meaning like *slowly* or *quickly*, as well as locatives denoting the place of eventuality (Michaelis, 1994; Mittwoch, 2008). Examples of this are found in (5) (after Mittwoch (2008)), where # and ? indicate that the sentence is not possible or sound odd on a resultative reading:

(5) (R-reading)

a. Berit **has arrived** in Paris. → Berit is in Paris.
b. ?The book **has been** wrongly **shelved** by one of the librarians.
c. #You**'ve corrected** the proofs too slowly.
d. #I **have peeled** the potatoes in the garden.

As a perfect, Modern Spanish *haber* + PTCP can have both existential and universal readings. The E-reading of the perfect is given in contexts with quantification adverbials like *a veces* 'sometimes', and with indefinite temporal specification like *antes* 'before' (see (6-a)); whereas the U-reading is clear in contexts with adverbs like *siempre* 'always' , as in (6-b). Resultative readings are clear in contexts, as in (6-c), where it is entailed that the state holds at the utterance time.

(6) a. *Marta **ha** **venido** muchas veces a Barcelona.*
Marta has come many times to Barcelona
"Marta has often come to Barcelona."

b. *Marta **ha** **vivido** siempre en Zaragoza.*
Marta has lived always in Zaragoza
"Marta has always lived in Zaragoza"

c. *Marta ya **ha** **llegado** a Barcelona.*
Marta already has arrived to Barcelona

"Marta has already arrived in Barcelona."

Before moving to explore the perfect contexts of use in Old Spanish participial constructions, a cautious note should be mentioned here about the similarities between resultative perfects and adjectival passives. As it will be seen in the following section, the distributional restrictions affecting perfects of result are similar to those of adjectival passives. Adjectival passives are possible with accomplishments and achievements, whereas they are mostly incompatible with IL-states and activities, and they show restrictions on the combinations with manner and other eventive modifiers, as well as spatio-temporal modifiers. However, there is a fundamental difference between the interpretations of both categories, concerning the interpretation of the entity which may be interpreted as the agent or cause of the underlying eventuality. While in adjectival passives this entity, in the absence of other context elements, may be considered identical to the grammatical subject (reflexive) or not (disjoint reference), as in (7) (after Gehrke (2012)), it is not possible to find this effect in resultative perfects, where the grammatical subject of the construction is always the same as the external argument entailed by the base predicate (only reflexive). This, which may seem quite obvious now, will be essential to my analysis of change and to disentangle the interpretation of some examples of participial constructions in Old Spanish.

(7) *Das Kind war schlampig gekämmt.*
the child was slopp(il)y combed
"The child was combed in a sloppy manner."
(i) Someone (else) (has) combed the child. *disjoint reference*
(ii) The child (has) combed himself/herself. *reflexive*

3.2.2 Perfects in Old Spanish

As in Modern Spanish, since the earliest centuries *haber* could appear with predicates yielding resultative, experiential or universal perfect readings. The perfect interpretation is clear in contexts in which base predicates are used intransitively. In these contexts it is not possible to interpret *haber* as a possessive verb, which obligatorily requires the presence of an object, interpreted as the possessed entity.

Examples of perfects with *haber* from the 13th century are found in (8), (9) and (10). Clear examples of existential perfects are found in (8), where *haber* is found with (intransitive) activities such as *andar* 'walk', *llorar* 'cry' or *trabajar* 'work'. Examples of *haber* with (intransitive) IL-states are found in (9). In the absence of temporal modification we cannot decide from the context whether the readings of the perfect periphrases are existential or universal, as states can yield both readings. Lastly, instances of *haber* + PTCP as in (10), where base predi-

cates are either achievements or accomplishments, thus yielding E- or R- perfect readings, are found in (10).

(8) (*Haber* with activities)

a. *de que* ***ouo andado*** *vn rrato por la çibdat fuese al*
of which had walked a while for the city went:REFL to the
palaçio
palace
"After he had walked across the city for a while, he went to the palace" (13th c., CZP)

b. *quando* ***ouo llorado*** *gran rato tomo la cabeça entre sus*
when had cried long while took the head between his
braços & besola muchas vezes.
arms and kissed her many times
"After she had cried for a long time, she held his head between her arms and kissed it many times." (13th c., VLT)

(9) (*Haber* with IL-states)

a. *[él]* ***hauie mereçido*** *obtener uictoria delos enemigos.*
[he] had deserved obtain victory from the enemies
"He had deserved to win over his enemies." (13th c., GCI)

b. *E quando cataljna dizia que tulio no* ***auja*** *tanto* ***valjdo*** *al*
and when Catalina said that Tulio not had so much cost to the
comun de Roma como el auja feyto
common of Rome as he had done
"And when Catalina said that Tulio had not served so much to the Romans as he had..." (13th c., LAT)

(10) (*Haber* with achievements and accomplishments)

a. *fallaron que* ***hauian*** *y* ***venido*** *los onze juezes*
found.3PL that had there come the eleven judges
"They found that eleven judges had come there." (13th c., BDS)

b. *entre las penyas el uido una carrasca. la qual se* ***auie***
between the crags he saw a holm oak the which REFL had
alli ***nasçido****.*
there born
"He saw a holm oak in the crags, which had emerged there." (13th c., GCI)

However, in the earliest centuries, *haber* could also combine with transitive states and activities yielding a perfect interpretation.These are contexts in which

the internal argument is a sentential complement. See examples illustrating these uses in (11), (12) and (13). In (13), where the base predicate is *saber* 'know', the complement to the participle is an infinitive which cannot be interpreted as the internal argument to *haber*, thus rendering the perfect interpretation possible. Besides, the use of iterative adverbial (11), or indefinite temporal adverbial *meses enantes* 'months before' (13) in combination with the present tense form of the auxiliary, clearly points to this reading. As can be seen from these examples, base predicates in these cases describe some type of cognitive or perception process. As argued by Benveniste (1968) (summarised in chapter 1), already in Late Latin *habeo* could combine with this type of predicates rendering a perfect interpretation.

(11) ***He*** ***pensado*** *muchas vezes. Muy yllustre y reuerendissimo señor*
have thought many times very illustrious and reverend lord
con que pudiesse en algo seruir a vuestra señoria
with that could in something serve to your lordship
Reuerendissima.
reverence
"I have often thought, my Lord, how I could please to Your Reverence." (16th c., M19)

(12) *todos sus caualleros ... fueron a enterrar el cuerpo del rey su*
all his knights ... went to bury the body of the king their
señor debaxo de monte caluarie segun ***hauedes oydo*** *ante*
lord under of mountain calvarie according have heard before
desto
of this
"All his knights went to bury their King's body under the Calvarie Mountain, as you have heard of this before." (13th c., VLT)

(13) *Muchos senyales fueron uistos de su muert ante que ella uiniesse si*
many signas were seen of his death before that she came if
el fuesse tan auisado bien sende ***aurie sopido aguardar*** *o* ***perçebir***
he were so warned well of that had known wait or perceive
.iij. o .iiijo. meses enantes.
iij or iiijo months before
"Many signals about his death were foreseen. If he had been warned he would have known many months before how to wait or watch." (13th c., GCI)

So far, apart from vocabulary and word order differences, all these examples are perfectly grammatical in Modern Spanish as well. However, in contrast with

Modern Spanish, in Old Spanish the perfect could also be instantiated by *ser* + PTCP. The periphrasis formed with *ser* is not grammatical in any of these contexts in Modern Spanish. The perfect interpretation is clear in contexts where base predicates are intransitive and thus the periphrasis cannot be interpreted as a verbal passive, as it obvious from the fact that passives are built from transitives. The fact that in these cases the grammatical subject of the construction is identical to the logical subject entailed by the base predicate further points to this interpretation.

In particular, E-perfects with *ser* are found in sentences where base predicates are activities, as in (14). Furthermore, examples such as those in (14-b) and (14-c) are not expected under Aranovich's hypothesis, summarised in chapter 1, that agentive verbs are expected to select *haber* (and not *ser*) to form the perfect since the earliest centuries.

(14) (*Ser* with activities)

a. *Esta dueña mj madre [...] nunca ouo por agujsado de me*
this lady my mother ... never had for prepared of me
castigar de palabra njn de fecho quando era pequeño njn despues
punish of word nor of fact when was little nor after
que ***fuy*** ***creçido*** *& loaua me todo quanto fazia quier*
that was grown and worshipped me all how much did if
fuese bueno quier fuese malo
were right if were wrong
"My mother never punished me when I was a child, and when I had grown up she worshipped all that I did; it did not matter whether it was right or wrong" (13th c., CZP)

b. *mas despues que ercoles ally* ***fue*** ***andudo*** *buscando la tierra*
but after that Hercules there was walked searching the ground
& aujsandola & semeJole muy buena. Et por ende
and warning-it and seemed-him very good and for that
poblo vna çibdat al pie de moncayo
inhabited a city to the foot of Moncayo
"But after Hercules had walked to find a place, it seemed to him a really good place, and for this reason he founded a city there, near Moncayo" (14th c., CRN)

c. *Y se echaua myo çid despues que* ***fue*** ***çenado*** *YVn*
and REFL lie Mio Cid after that was had dinner and a
suenol priso dulçe tan bien se adurmjo
dream-him took sweet so well REFL slept
"And Mio Cid lie down after he had had dinner. And he fell deeply asleep" (12th c., CID)

Cases of *ser* with intransitive accomplishments and achievements, thus yielding R- or E-perfects are found in Spanish since the earliest centuries as well. See examples in (15). From the context in (15-a) and (15-b), it is not possible to know whether the it holds a R- or E-reading; in (15-c), the presence of the indefinite temporal adverbial *antes* 'before' points directly to an E-reading.

(15) (*Ser* with accomplishments and achievements)

a. *El infante **fue** **uenido**; por las armas prender.*
the infant was come to the weapons take
"The infant had come in order to take the weapons." (13th c., ALX)

b. *El princepe Aristomoles; en Egipto **fue** **nado**.*
the prince Aristomoles in Egypt was born.
"The prince Aristomoles has been born in Egypt." (13th c., ALX)

c. *los çagueros se tornaron entala Ribera del Rio de*
the hinds REFL went back in that bank of the river from
*do ellos **eran** **partidos** mas antes*
where they were gone more before
"The hinds went back to the bank of the river, where the had left long time before." (13th c., GCI)

IL-states with *ser*, yielding universal readings, are not found with this copula in the corpus. Examples in sentences such as (16) and (17), where the verb used is written the same as the stative predicate *parecer* 'seem', it is nevertheless the verb *aparecer* 'appear'. Both were homonyms in Old Spanish. The latter interpretation of this verb in these examples is clear from the use of prepositional phrases (*en aquel logar* 'in that place' in (16), *all orizon* 'above the horizon' in (17)) describing the location where the event of coming into existence takes place.

(16) *Et quando fuere la longura del Sol del punto del mudamiento*
and when were the lenght of the sun of the point of the changing
estiual entre los dos grados que son oppositos destos dos grados
estival between the two grades that are opposed of these two grades
*sobredichos; non **sera** **parescida** en aquel logar.*
above mentioned not be appeared in that place
"And when the size of the sun is at the point of the change of season, between the two degrees which are opposite to the degrees above mentioned, then it will not have appeared in that place." (13th c., ALB)

(17) *Et otrossi si **fuere** este cerco **parescido** all orizon quando sube el*
and then if were this ring appeared to the horizon when rises the

Sol. & quando se pone. sera el punto de oriente o de
sun and when REFL goes down be the point of Orient or of
occidente
Occident
"And if that ring had appeared in the horizon when the sun rises, when it goes down it will be the point of Orient or Occident" (13th c., ALB)

On the other hand, both *ser* and *haber* plus intransitives appear in **modal or irrealis contexts** in Old Spanish. See examples of *ser* + PTCP in these contexts in (18). Examples with *haber* in similar contexts, which are grammatical in Modern Spanish, are in (19).

(18) (*Ser* + PTCP in modal or irrealis contexts)

a. *si aquella infortuna **fuere** **sallida** del ascendente a la*
if that infortune was.SBJV gone out of the ascendent to the
.xija. casa; significa que fue cercada.
XIJ house means that was surrounded
"If that infortune had come from ascendent in the xija house, it means that it was surrounded." (13th c. JUZ)

b. *E si los xpistianos **fuessen** **llegados** antes que los Turcos a*
and if the christians were.SBJV arrived before that the Turcs to
aquell logar por fuerça se tornaran los moros a Egipto
that place for force REFL go back the moors to Egypt
"And if the Christians had arrived before than the Turcs to that place, the moors would have gone back to Egypt" (13th c. ULT)

c. *Mas si **serja** **venjdo** por nos hazer a nos otros otro tanto*
but if was.COND come for us make to us others other much
yo me pondre alo menos en buen rrecabdo
I me put to the less in good place
"But if he had come to make us other I will put myself in a safe place" (15th c., MEL)

(19) (*Haber* + PTCP in modal or irrealis contexts)

a. *si **ouieran** **llegado** antes & **andado** mas que anduuieron*
if had.SBJV arrived before and walked more than walked.3PL
ouiera ganado ropin el reyno de armenia
had won Ropin the kingdom of Armenia
"If they had come before and walked longer that they did, Ropin would have won the Kingdom of Armenia" (13th c., VLT)

b. *el hi huuies estado preso & muerto con el*
he there had.SBJV been imprisoned and died with the

Romanient sino ***huuiessen foydo*** *en una montanya*
Roman unless had.SBJV escaped in a mountain
"He would have been imprisoned and killed with Romanient unless they had escaped to a mountain" (13th c., GCI)

Interestingly, this contrasts with other diachronic language varieties like Old English, where it has long been known that perfects with *be* are strongly restricted in modal contexts (Fridén, 1948; Rydén and Brorström, 1987; Kytö, 1997; Traugott, 1972; Johannisson, 1958; McFadden and Alexiadou, 2005, 2010, among others). In particular, McFadden and Alexiadou (2010) argue on the basis of a large-scale corpus that the periphrasis with *be* was restricted to a (stative) resultative interpretation during late Middle English (from 1150 to 1500) and Early Modern English (from 1500 to 1710), as it can only very rarely be found in past counterfactuals or with durative and iterative adverbials. By **past counterfactuals** these authors refer to clauses that convey that the proposition being considered was contrary to fact at a particular time in the past, as in (20), after McFadden and Alexiadou (2010, p. 395)

(20) a. *and if they had come sooner, they could haue holpen them.*
and if they had come sooner they could have helped them
"And if they had come sooner, they could have helped them." (Giff,G3V.246)

b. *he had never come to himself ... if he had not met with this*
he had never come to himself ... if he had not met with this
allay
distraction
"He would never have come to himself ... if he had not met with this distraction." (Behn,189.165)

Lastly, there is a context in which, as in the case of IL-states, it can be observed a categorical contrast in the use of *ser* or *haber* as perfect auxiliaries in Old Spanish. Perfects of progressives, yielding U-readings, are only attested with *haber*; there are no examples in the corpus of perfect of progressives with *ser*. See some examples in (21) and (22).

(21) *desy sopo commo aquel su cauallero ferrand antolines en cuya*
of if knew how that his knight Ferrand Antolines in which
figura el otro ***auja estado peleando*** *durante la batalla*
figure the other had been.LOC fighting during the battle
"About whether he knew how his knight Ferrand Antolines, against whom the other had been fighting during the battle." (13th c., CD2)

(22) *Acabada la furia y virtud de la yerua, boluian en si, y*
finished the fury and virtue of the grass, came in themselves and
dezian ***auer estado comunicando*** *el negocio con el demonio*
said have been.LOC communicating the issue with the demon
"Once the fury and virtue of the grass ended, they came back to be themselves, and they said that they had been talking with the demon" (16th c., IND)

So far, we have seen instances of perfects in Old Spanish with *ser* and with *haber*, as in Modern Spanish. But, crucially, in Old Spanish other auxiliaries, such as *estar* 'be.LOC' or *tener* 'have.POSS', show perfect uses as well. In contrast with Modern Spanish, where periphrases with *estar* and *tener* are not grammatical in perfect contexts, since the earliest centuries both *estar* and *tener* could appear with predicates in which this is the only possible interpretation.

Perfect uses of *estar* + PTCP are clear in cases where base predicates are intransitive accomplishments or achievements describing motion or change of location. Some examples illustrating these uses are in (23), and (24). In these examples base predicates are used intransitively and the subject is an animate entity, which makes a dynamic interpretation of motion more available. Furthermore, in some of these cases the periphrasis is modified by directional prepositional phrases, which are only possible if predicates are interpreted non-statively (see *al muro* 'to the wall' (23), and *a otra part* in (24)). Examples of this kind are attested in the corpus until the 16th century. In the absence of any additional contextual information, such as temporal adverbials, we cannot really know whether these are instances of E- or R- perfects.

(23) *Quando vieron los turcos aquellos vellacos que* ***estauan*** *ya*
when saw the turcs those means that were.loc already
llegados *al muro vinieron de todas partes & traxeron huego*
come to the wall came of every part and brought fire
grecisco encendido que parecia sangre tanto era bermejo
greek reddened that seemed blood so was red
"When the Turks saw that mean people had already come to the wall, they came from all parts and brought fire which look red as blood" (13th c., VLT)

(24) *Et llegando a la oriella del lago non ueyendo el aun. ca a otra*
and arriving to the bank of the lake not seeing him yet as to other
part ***estaua*** ***tornado***. *echo luego el manto aluenne*
part was.LOC come back threw then the cloak far
"And upon arriving to the bank of the lake and not seeing him yet, as he

had gone back to the other side, he through the cloak far away" (14th c., G2K)

Perfect uses of *tener*, which are not possible in Modern Spanish, are found with predicates describing some type of communication or cognition state or event. These cases are similar to those we saw before, in (11), (12) and (13), with *haber* as the perfect auxiliary. The absence of agreement in some of these cases points further to the perfect interpretation of these constructions, as this is a morphological characteristic of perfect participles in Modern Spanish. See examples of *tener* with predicates describing some communication process in (25) and (26), and with predicates describing some state of knowledge in (27). Sentential complements in these cases, which are formed by infinitives, can hardly be interpreted as complements of *tener* in its possessive interpretation.

(25) *tristan & la rreyna* ***tenjan*** ***fablado*** *de dormjr en vno aquella*
Tristan and the queen had.POSS talked of sleep in one that
noche
night
"Tristan and the Queen had talked about sleeping together during that night" (14th c., CAP)

(26) *Por que no faga como el que se desuia del camino que*
for that not do as the that REFL divert of the way that
quanto mas anda: tanto mas se alexa del termino:
how much more walks so much more REFL moves away of the place
a donde ***tenia*** ***pensado*** *de yr.*
to where had.POSS thought of go
"So that he did not do as the one who the longer he walks, the more he loses his way and moves away from the place where he had decided to go." (15th c., CAP)

(27) *Respuesta del rey. POr bien aconseiado me tuuiera de vosotros*
answer of the king for well advised me had.POSS of you
sino ***tuuiese*** ***sabido*** *ser tan devido vengar las desonrras como*
unless had.POSS known be so urgent revenge the dishonor as
perdonar las culpas.
forgive the blame
"Answer of the King: I would consider myself well advised from you unless I had known that one must take revenge for dishonor as well as forgive the sins." (15th c., CAR)

Other transitives can combine with *tener* to yield perfects. For example, in

(28) the absence of agreement between *cuevas* 'caves' and the participle *fecho* 'done' suggest this interpretation. In (29) the adverb *bastantemente* needs to be interpreted with respect to some agent, which is not possible under a stative interpretation.

(28) *et los que **tenian** **fecho** cueuas en los cuestos impliensse*
and the who had.POSS done:M.SG caves:F.PL in the hills fill:REFL
de agua
of water
"And those who had done caves in the hills, filled in with water" (13th c., GC3)

(29) *el verdadero Carpesio tan celebrado por los antiguos ni es a*
the truth Carpesio so celebrated for the ancient nor is to
*nuestro proposito tratar lo, ni faltan autores graues, que lo **tengan***
our purpose deal it nor lack authors serious that it have
*bastantemente **demostrado**.*
greatly proved
"It is not our intention to deal with the truth Carpesio, which was so celebrated by the ancient people, as there are many serious authors which have greatly proved it." (16th c., IND)

The last attested examples of *tener* with perfect uses in the corpus are found in the 16th century. After this century, occurrences of this type are found with a very limited number of predicates. There are some examples, as in (30) and (31), in the 19th century but they only appear in texts written by Blasco Ibañez, and therefore cannot be considered representative of the system of that time.

(30) ***tengo** **salido** muchas veces diciendo: No faltaré, no*
have.POSS gone out many times saying not will be missing not
faltaré.
will be missing
"I have often ended up by saying: I will go, I will!" (18th c., GHPQ)

(31) *Alguna vez se le **tiene** **escapada** ésta y otras*
some time REFL him has.POSS escaped this and other
exclamaciones semejantes: ¡Cómo me carga este chiquillo!
exclamations similar how me burden this child
"And in some occasions he says this or similar exclamations: What a burden is this child to me!" (18th c., GHPQ)

To sum up, on the basis of the facts just presented, we can conclude that experi-

ential and resultative perfect readings could be formed since the earliest centuries using *haber*, *ser*, *tener* or *estar*. The absence of *ser*, *estar* and *tener* with progressives in the corpus suggests that universal readings with *ser* may have not been possible in Old Spanish. On the other hand, the contexts in which perfects formed with *estar* and *tener* appear suggest that they might have been interpreted only as resultatives in Old Spanish.

We have also seen that the distribution of perfect periphrases in Old Spanish contrasts with the contexts of appearance of cognate constructions in earliest stages of other languages like English, where auxiliary selection in the perfect has also been lost (McFadden and Alexiadou, 2010). In contrast with the perfect with *have*, in Earlier English the use of perfect *be* was restricted to a stative resultative (perfect of result) interpretation, whereas that with *have* developed a wider range of uses, crucially including the experiential perfect in addition to resultatives.

3.3 Passive contexts of participial constructions

3.3.1 Basic facts about verbal and adjectival passives

Moving now to passives, there are two well-known facts about the distribution of adjectival and verbal passives which clearly differentiate between both categories (Kratzer, 2000). The first fact concerns event-related modification and the second the types of predicates that can form adjectival and verbal passives.

Examples in German are used to illustrate these facts, as this language has two distinct morphological expressions for these interpretations and, on the other hand, most seminal papers about the semantics of these categories are based on evidence from this language.

Firstly, adjectival passives cannot generally combine with event-related modifiers whereas verbal passives can. Event-related modification includes instrumentals, manner and spatio-temporal modifiers and *by*-phrases. As can be seen from examples in (33), after Rapp (1996) and Gehrke (to appear), adjectival passives are not possible in contexts where the agent is expressed by an oblique prepositional phrase, or with instrumentals and other adverbials that modify eventive uses of verbs, and which are necessarily interpreted with respect to some agent, as in (32).

(32) *Der Mülleimer* ***ist*** *(*von meiner Nichte/ *langsam/ *genüsslich/ *mit der Heugabel)* ***geleert***.
the rubbish bin is by my niece slowly pleasurably with the hay fork emptied

On the other hand, only verbal passives can take prepositional phrases describing

the place or time in which the eventuality takes place, whereas adjectival passives cannot. See examples in (33-a) and (33-b). Spatio-temporal modification may appear with adjectival passives as well, but it is restricted to cases where it describes the location of the result state, but not where the previous (entailed) event took place. For example, in (33-b) in the verbal passive with *werden* the locative prepositional phrase *in der Garage* is interpreted as the spatial location where the actual inflation event takes place. In contrast, the adjectival passive interpretation with *sein* (*sind*) and the locative prepositional phrase is only possible under a reading in which the tires lie inflated and they stay in the garage.

(33) a. *Der Computer* ***wurde / *ist*** *vor drei Tagen* ***repariert.***
the computer is before three days repaired
b. *Die Reifen* ***wurden / ??sind*** *in der Garage* ***aufgepumpt.***
the tires are in the garage inflated

As for predicate restrictions, verbal passives can combine with most types of transitives whereas adjectival passives are built up from a more restricted range of predicates, including some unacussatives as well. From the aspectual point of view, adjectival passives can build only from verbs encoding some consequent state in their lexical semantics such as accomplishments and achievements. Transitive activities and IL-states cannot build adjectival passives but they can build verbal passives, probably because they do not include a consequent state component in their meaning (or BECOME in Levin and Rappaport (1994), and subsequent work) (34).

(34) *Die Katze* ***wurde/*ist gestreichelt.***
the cat is petted

However, it has been observed that particular event-related modifiers are allowed with adjectival passives in some contexts (Kratzer, 2000; Gehrke, to appear). Despite the fact that the construction is stative in nature, it can combine with some event-related modifiers, which are necessarily interpreted with respect to some agent. See some examples of adjectival passives with instrumentals and other agentive adverbials from German in (35).

(35) a. *Der Brief* ***ist*** mit roter Tinte ***geschrieben.***
the letter is with red ink written
b. *Das Haar* ***war*** *schlämpig* ***gekämmt.***
the hair was sloppily combed

Despite appearances, event-related modifiers are still different when they appear with adjectival and verbal passives. Event-related modifiers appearing with stative passives tend to have a more generic character than those combining with ver-

bal passives, which are often more specific (Schlücker, 2005; Gehrke, to appear; Gehrke and Sánchez-Marco, 2012). As argued in Gehrke and Sánchez-Marco (2012) based on quantitative evidence from a Modern Spanish corpus, *by*-phrases with adjectival passives appear more frequently with bare nouns and indefinite determiners, thus yielding a generic interpretation, whereas those *by*-phrases with definite determiners, pronouns and proper names appear more frequently with verbal passives. From now on I will refer to *by*-phrases of the former type as **generic** and those which tend to combine with verbal passives as **specific**. See some examples of adjectival and verbal passives with these types of *by*-phrases in Modern Spanish in (36) and (37).

(36) (Generic *por*-phrases)

a. *Esta guerra **está / ??es dirigida** por militares incompetentes.*
this war is.LOC is leaded by soldiers incompetent
"This war is leaded by incompetent soldiers."

b. *El piso que ahora habito no está arreglado por obreros, sino*
the flat that now live in not is.LOC repaired by workers but
*por personas que se hacen pasar por obreros. **Está / ??Es***
by people that REFL pass for workers is.LOC is
***arreglado** por un piloto*
repaired by a pilot
"The flat in which I am living now is not fixed by workers, but by people who pretend to be workers. It is fixed by a pilot"

(37) (Specific *por*-phrases)

a. *La casa ??**estuvo / fue pintada** por Juan / él.*
the house was.LOC was painted by Juan him
"The house was painted by Juan/him"

b. *La cultura urbana perdió el respeto y el correlato de la*
the culture urban lost the respect and the correlate of the
*Naturaleza cuando la producción agrícola ??**estuvo / fue***
nature when the production agricultural was.LOC was
***sustituida** por la producción industrial.*
replaced by the production industrial
"The urban culture lost the respect and correlate with nature when agricultural production was replaced by industrial production" (20th c., a24)

3.3.2 Passives in Old Spanish

On the basis of the basic facts about the verbal and adjectival passive categories, as just presented, in the following I explore the passive contexts of use of the periphrases in Old Spanish.

In Old Spanish passives were expressed by *ser* + PTCP and *estar* + PTCP. Apart from the perfect uses of participial constructions with *ser* and *estar*, shown in the previous section, both auxiliaries could appear in other contexts, specifically verbal passive contexts, presented in Section 3.3.2.1, and adjectival passive contexts, in Section 3.3.2.2. In this respect, there is a difference as regards the expression of the perfect in Old Spanish, which could be expressed by periphrases formed with all four auxiliaries.

3.3.2.1 Verbal passive contexts

In contrast with Modern Spanish, when verbal passives are expressed only by *ser* + PTCP, in Old Spanish both *ser* and *estar* appear in contexts in which they are clearly interpreted as verbal passives. These include contexts of use of *ser* and *estar* in combination with certain types of verbs and with some event-related modifiers, in which neither an adjectival passive nor a perfect interpretation is possible.

In Old Spanish *estar* could combine with transitive IL-states like *amar* 'love' and *creer* 'believe'. In these contexts an adjectival passive interpretation is not possible, as IL-states encode a consequent state in their meaning, which is required to form adjectival passives. In Modern Spanish, *estar* is not possible in these contexts. Examples in (38) illustrate this point.

(38) (*Estar* with transitive IL-states)

a. *Contra esto dize el otro que mas val* ***estar*** ***amado*** *que*
against this says the other that more is worth be.LOC loved than
temjdo
feared
"The other says against this idea that it is better to be loved than feared" (13th c., LAT)

b. *la dicha aliança porel scilençio de tantos tiempos*
the above mentioned alliance for the silence of many times
.iiijo. ***ha estado tenida*** *en tal guisa que ella es estada*
iiijo has been.LOC had.POSS in such guise that she is been.LOC
aprouada
aproved
"The above mentioned alliance for the silence of so many times has

been hold in such a guise that it has been finally approved" (13th c., GCI)

Secondly, in the earliest centuries participial periphrases with *estar* are found with event-related modifiers such as instrumental manner adverbials and *por* 'by' phrases which are necessarily interpreted with respect to some agent. In Modern Spanish, however, there are rarely found cases of *estar* + PTCP with this type of modifiers. Examples of *estar* + PTCP with manner instrumentals illustrating this point are in (39); in (39-a) the modifier is *conlas manos & armas* 'with his hands and weapons' and in (39-b) is *con armas* 'with weapons'.

(39) (*Estar* + PTCP with instrumentals)

a. *es asaber quelas anjmas de los justos todos tienpos* ***estan***
is to know that the souls of the fair all time are.LOC
guardadas *conlas manos & armas del su muy alto padre*
kept with the hand and arms of the his very high father
"It should be noted that the souls of the fair are kept forever with the hand and weapons of their Father." (15th c., DON)

b. *mas ama la sangre mia que ninguna otra cosa del mundo.*
more loves the blood mine that no other thing of the world
ya ha .v. meses que yo companyon & amigo del pueblo
already has v months that I colleague and friend of the people
romano ***so estado*** ***assitiado*** *con armas.*
roman am been.LOC besieged with weapons
"He loves my blood more than any other thing in the world. 5 months ago already it is been since I was besieged with weapons." (13th c., GCI)

Similar examples from Old Spanish with *ser* in combination with IL-states and instrumentals, therefore grammatical in Modern Spanish, are found in (40-a) and (41), respectively. Transitive activities expressing verbal passives can be found with *ser*, as in (40-b); however, similar examples with *estar* are more rarely found in the corpus.

(40) (*Ser* with transitive IL-states and activities)

a. *Fue tenido por ombre bueno; & de seso acabado.* ***fue***
was had.POSS for man good and of brains finished was
dalli adelantre; ***temido*** *&* ***amado***.
of there onwards feared and loved
"He was considered a good man. Since then, he was feared and loved." (13th c., ALX)

b. *Cuemo sy **fuessen** siempre; enello **usadas**.*
as if were always in that used
"As if they were always used to do that" (13th c., ALX)

(41) (*Ser* with instrumentals)

*Sea puesto que dos varones **sean feridos** con espada o con*
be put that two men be wounded with sword or with
ssemejante cuchillo en vna mesma ora
similar knife in a same time

"It should be written that two men have to be wounded with a sword or knife at the same time" (15th c., CMN)

In Old Spanish there are also other contexts that suggest a verbal passive interpretation. These are, firstly, contexts in which participial periphrases appear in combination with specific *por*-phrases, such as those with proper nouns, pronouns, and definite or specified noun phrases. As mentioned before, this type of phrases tend to appear only with verbal passives. Examples from Old Spanish with *estar*, which are not grammatical in Modern Spanish are found in (42).

(42) (*Estar* + PTCP with specific *por*-phrases)

a. *todo aquesto fizo en memoria dela filla. la qual cosa no*
all this made in memory of the daughter the which thing not
***era** encara **estada** **começada** por ninguno*
was still been.LOC begun by nobody
"He did all this in memory of his daughter. All that had not been started by anybody." (13th c., GCI)

b. *E mandamos que los que agora **estan** **puestos** por ellos non*
and send that the that now are.LOC put by them not
vsen de los dichos ofiçios en el dicho adelantamjento
use of the mentioned trades in the mentioned territory
nin sean obedesçidos nin avidos por alcaldes
nor be obeyed not had by mayors
"And we ordered that those who now are put by them should not use those trades in that terroritory neither should be obeyed nor considered mayors." (15th c., MTV)

Temporal frame adverbials and spatial modifiers that pick out the location of the event that brought about the state can also appear with *estar* + PTCP in Old Spanish, whereas they are agrammatical in Modern Spanish. See examples in (43).

(43) (*Estar* + PTCP with spatio-temporal modifiers)

E esto fazian continuadamente porque en el menester & rrebate
and this made continuously because in the occupation and refute
de las batallas subiesen apriesa en sus cauallos los que asy ***estauan***
of the battles rised quickly in their horses the that so were.LOC
vsados *en el tienpo de la paz.*
used in the time of the peace

"And they did that continuously so that in the battle they quickly ride in their horses, the ones which were used during the peaceful time." (15th c., LVC)

Similar examples from Old Spanish with *ser* + PTCP, which are perfectly possible in Modern Spanish, are in (44), where *ser* + PTCP appears with specific *por*-phrases, and in (45), where it appears with spatio-temporal modifiers.

(44) (*Ser* + PTCP with specific *por*-phrases)

a. *tantos dannos podria adozir a otros por tal iura que*
many damages could put forward to others for such act that
nunqua podrian ***seer emendados por el****.*
never could be amended by him
"And there were many damages produced by the others which could never be amended by him." (13th c., FAR)

b. *ET apres desque el barbaro* ***fue muerto por Çipion*** *las dos*
and after since the barbarian was killed by Cipion the two
huestes se ayuntaron ensemble
armies REFL joined together
"And after the Barbarian was killed by Cipion the two armies joined." (13th c., GCI)

(45) (*Ser* + PTCP with spatio-temporal modifiers)

cosa que se a fecha contra qual quiera destas non vale njn
thing that REFL has done against which wants of these not worth nor
deue ***ser tenjda*** *njn* ***guardada*** *& enel tienpo que se falla*
must be had nor kept and in the time that REFL finds

"There is a thing which is done against any of these which is not worth neither it must be hold nor kept in the time when it is found." (13th c., CAX)

3.3.2.2 Adjectival passive contexts

In Modern Spanish, the participial periphrasis with *estar* is the only morphological instantiation of adjectival passives. In contrast, in the earliest centuries participial constructions formed both with *ser* and with *estar* could be interpreted as adjectival passives.

Adjectival passive contexts are those in which the copula combines with participles denoting result states, but which are nevertheless not compatible or rarer with verbal passives, such as those in which base predicates are object experiencer psychological verbs like *maravillar* 'wonder' or locative verbs like *poner* 'put'. Sentences in (46) and (47) illustrate these adjectival passive contexts with *ser* in Old Spanish. All these examples are not grammatical in Modern Spanish.

(46) (*Ser* with object experiencer psych predicates)

a. *Ya dezien los de greçia; que* ***eran enoiados****. Que la yra de*
already said the of Greece that were annoyed that the anger of
dios; gelos auie deparados.
god REFL them had held
"The Greeks already said that they were annoyed; that the anger of God has been brought to them." (13th c., ALX)

b. *E por que tenja que njngund omne non deuja fazer mas*
and for that had.POSS that no man not should do more
por guardar el fecho del Rey que aquel don nuño que ***era***
for keep the fact of the king that that mr Nuño that was
maraujllado *desto quele dezjan*
amazed of this that him told
"And because he thought that no man should make anything else to keep the king's fact. Mr Nuño was amazed at this that was told to him." (13th c., CAX)

Furthermore, when these copulas are combined with locative verbs, as in (47), there is in most cases other contextual information that suggests that these periphrases hold a stative interpretation. Specifically, these are cases where the periphrasis is modified by an adverbial describing the place where the entity described by the internal argument (coded as the grammatical subject) is placed; see *en su casa* 'in his house' in (47-a) and *en este libro* 'in this book' in (47-b).

(47) (*Ser* with locative verbs)

a. *Esto es por fuero de omne que dize que* ***es encerrado*** *en su casa*
this is for law of man that says that is locked up in his house
"This is by law of man who is locked up in his house" (13th c., CAS)

b. *las razones & las palabras. & las leyes que* ***son*** ***contenidas*** *en*
the reasons and the words and the laws that are contained in
este libro. bien queremos que de aqui en adelantre que non sean
this book well want that of here in forward that not be
husadas las leyes romanas.
used the laws roman
"The reasons and words and laws which are contained in this book; we really want that from now on the Roman laws are not used" (13th c., FJZ)

Furthermore, in Old Spanish *ser* + PTCP could appear in contexts with generic *por*-phrases, which suggest an adjectival passive interpretation. This is illustrated in (48), where *por*-phrases appear with noun phrases headed by definite determiners with nouns describing abstract entities (48-c), or plural noun phrases with a generic character, as in (48-a) and (48-b).

(48) (*Ser* + PTCP with generic *por*-phrases)

a. *Et assi el emperado leo releuo el imperio romano el qual*
and so the emperor Leo relieved the empire roman the which
era ***opremido*** *por muchas miserias*
was oppressed by many misery
"And so the emperor Leo relieved the Roman empire, which was oppressed by many misery" (13th c., GCI)

b. *E otrosi a los que començaron el camjno para yr enesta*
and therefore to the that begun the path to go in this
*rromeria con voluntad dela conplir & **fueron** **enbargados***
procession with will of it accomplish and were overcome
por enfermedades & por otros enbargos
by diseases and by other seizures
"And so to those who begun the path for this procession with the will to finish, they were overcome by diseases and other seizures." (13th c., CZP)

c. *Todos los rreligiosos digo que todos estos entredichos non*
all the religious say that all these mentioned not
aguardaren bien sin todo enganno encobierto &
keep good without all deceive covered and
descobierto qual que ssea ssegund ***es*** ***ordenado*** *por el derecho*
uncovered which that be according is ordere by the law
"All the religious people, I say that all this above mentioned, should

not go on well without all the deceive being uncovered, whichever it is, so it is ordered by the law" (14th c., CF1)

Similar adjectival passive contexts of use with *estar*, all of them grammatical in Modern Spanish, are in (49), (50), and (51):

(49) (*Estar* with object experiencer psych predicates)

a. *Et ellos tanto* ***estauan*** ***enoiados*** *que lo touieron por bien Et*
and they so were.LOC annoyed that it had.POSS for good and
dixieron que se querien yr et nunca tan bon dia vieran
said that REFL wanted go and never so good day saw
"And they were so annoyed that they took it well. And they said that they wanted to leave and that they had never seen such a good day" (13th c., EE2)

b. *como uido a todos* ***estar*** ***marauellados*** *si les começo a*
as saw to all be.LOC amazed yes them begun to
declarar por tal manera las palauras.
declare for such manner the words
"As he saw that everyone was amazed, he started to declare his words in the following way..." (13th c., GCI)

(50) (*Estar* with locative verbs)

a. *commo clauo que* ***esta*** ***fincado*** *en alguna tabla*
as nail which be.LOC stuck in some plank
"...as a nail which is stuck in some plank." (13th c., CC2)

b. *E pues commo asy* ***estaran*** *aqui syenpre* ***ençerrados***
and then as this will.be.LOC here always locked-up
"And therefore they will stay here locked up forever." (13th c., CZP)

(51) (*Estar* + PTCP with generic *por*-phrases)

a. *et todos los que alli eran plegados dixieron que comoquier*
and all the that there were joined said that however
quela tierra ***staua*** ***enoyada*** *por las muytos peitos que*
that the land was.LOC annoyed by the many conflicts that
auian peitado por las guerras que auia auido
had been by the wars that had had
"And all that they were there together they said that the land was annoyed by all the conflicts as there have been there, by the wars which had taken place there" (13th c., GC3)

b. *que assi como los nuestros coraçones* ***estan*** ***crebantados*** *por*
that so as the our hearts are.LOC disturbed by

la soberuia daquellos assi nos allegremos nos . & nos
the arrogance of those so us cheer up us and us enjoy
gozemos de la nuestra humildad.
of the our humbleness
"..that in the same way as our hears are disturbed by the arrogance of others, so we should be happy and enjoy our humbleness" (13th c., GE4)

Lastly, in Old Spanish, participles with both *estar* and *ser* could be modified by adverbial *todavía* or *aún* 'still', expressing an adjectival passive interpretation. In Modern Spanish, neither a perfect nor a verbal passive interpretation is possible in these contexts. As has been noted in previous studies (Kratzer, 2000), this adverbial modifies states that are in principle reversible or transitory (so-called *target states* in Kratzer's terms); whereas it is incompatible with participles denoting states that hold forever after (*resultant states*).

(52) (*Estar* + PTCP and *ser* + PTCP with adverbial *todavía* 'still')

a. *Mas ieroboam fijo de nabath* ***estando*** *en egipto avn* ***foydo***.
but Ieroboam son of Nabath being.LOC in Egypt still escaped.
"But Ieroboam, son of Nabath still being escaped in Egypt" (13th c., IJ8)

b. *Capitulo que muestra commo* ***somos*** *avn* ***obligados*** *por otras*
chapter that shows how are still forced for other
rrazones. Deuemos segunda mente ser jncljnados A orar parando
reasons must second -ly be inclined to pray stopping
mjentes al enxenplo delos santos padres pasados.
minds to the example of the sain fathers past
"Chapter that shows how we are still forced for other reasons. Secondly, we must be inclined in order to pray, paying attention to the example of earlier saint fathers." (15th c., DON)

3.4 Stative possessive contexts of participial constructions

Apart from perfects, in the earliest centuries participial constructions formed with *haber* and *tener* could appear in contexts with stative possessive readings.

Modern Spanish *tener*, as English *have* can be used in some contexts describing a possessive relation, which may be inalienable or alienable depending on the relation between the subject of *have* (possessor) and the internal argument (pos-

sessee). See examples from English in (53), (53-a) and (53-b), after Harley (1998). In these cases there could a secondary predicate, merely describing a property of the internal argument, without eventive implications, as in (53-c). In these cases the past participle is a secondary predicate that describes a situation related to the internal argument, which is the entity possessed by the subject of the auxiliary.

(53) (Possessive *have*)

a. Getafix **had** a golden sickle. (possessive)
b. The oak tree **has** a nest in it. (possessive locative)
c. Getafix **had** a handmade sickle.

Other possible uses of English *have* are experiencer and causative uses, as has been explored, among others, by Belvin and den Dikken (1997); Déchaine *et al.* (1994); Ritter and Rosen (1997); Harley (1998). It is useful to consider this here, as similar interpretations might have been possible in Old Spanish, as we will show in what follows. These uses are not available to *haber* + PTCP nor *tener* + PTCP in Modern Spanish.

In experiencer uses the subject of the auxiliary is affected by situation expressed by the complement, which in the present case study is a past participle, but which could also be a gerund, infinitive, or an oblique prepositional phrase. Experiencer uses describe an event or situation which adversely affects the subject. See some examples in (54), after McIntyre (2006).

(54) (Experiencer *have*)

a. He **had** students **walking out of his lecture**.
b. John **had** his camera **in the water**.
c. John **had** his camera **smashed**.
d. He **had** people **destroy his car**.

In contrast, in causative readings the subject of the auxiliary rather causes the situation in the complement to come about, as in (55). In causative contexts the subject of *have* is the causee of the eventuality denoted by the complement, which could be a participle, but also a prepositional phrase, an adjective, a bare infinitive or a progressive. See some examples from English in (55) (after Harley (1998)):

(55) (Causative *have*)

a. Asterix **had** Obelix **deliver** a menhir to Getafix.
b. Asterix **had** Obelix **locked** in his hut.
c. Asterix **had** Obelix **running** errands for him.
d. Asterix **had** Obelix **red** in the face.
e. Asterix **had** Obelix **on the leftmost horse**.

The main difference between experiencer and causative uses is whether they fulfill the so-called **link requirement** (after Belvin and den Dikken (1997)), or not. The link requirement requires that the grammatical subject of the auxiliary is coindexed with a pronoun or variable somewhere in the complement. Experiencer readings are only possible if this requirement is fulfilled, whereas causative readings are preferred if this requirement is unfulfilled (although not impossible otherwise).

In what follows I will illustrate with examples stative possessive contexts, as well as experiencer and causative contexts of use of *haber* + PTCP and *tener* + PTCP in Old Spanish.

3.4.1 Stative possessive contexts

In stative possessive contexts the participle describes a property of the internal argument, which may be in alienable or inalienable relation with the subject of *haber* or *tener*. Both *tener* + PTCP and *haber* + PTCP could appear in these contexts. In Modern Spanish only *tener* + PTCP is possible in these contexts.

In Old Spanish, stative possessive readings of *haber* + PTCP, which are not possible in Modern Spanish, are obvious in contexts in which the entity coded as subject is in a possessive relation with the object. See some examples illustrating these uses in (56), (57) and (58). Example (56) describes an alienable possession relation, the external argument must be animate, since inanimates cannot own things. Inalienable possession uses of *haber* in Old Spanish are evident in contexts, as in (57) and (58), in which the complement is an inalienable possession or any kind of relational nominal. The external argument need not be animate. It is however interesting to observe that in Old Spanish, inanimate subjects with *haber* are rarely found. In most cases the possessed entity and possessor hold a part-whole relationship, which is very frequently a part of the body.

(56) *Et assy como las naues se aplegaron Guilchalco qui* ***auie*** *su*
and so like the ships REFL came closer Guilchalco who had his
nau bien ***armada*** *de buenos caualleros començo a combatir fuertment*
ship well armed of good knights begun to fight strongly
"And so like the ships came closer, Guilchalco who had his ship well armed with good knights started to fight strongly" (14th c., CQ1)

(57) *lo primero les contesçe por que [los viejos]* ***han*** *los cuerpos*
the first them happen because the elderly have the bodies
ya muy ***enfriados*** *& falleçe enellos la calentura natural la*
already very coldened and dies in them the heat natural the

qual faze alos omnes auer cobdiçias destenpradas de luxeria
which makes to the men have greed harsch of luxury
“” (13th c., CD2)

(58) *pregonaron por las tiendas [...] que se armassen lo mejor que*
announced through the shops ... that REFL armed the best that
pudiessen & viniessen a combatir la cibdad aderredor muy
could that came to fight the city around very
esforçadamen te: que mucho ***auian*** *los coraçones* ***encendidos*** *&*
hardly that much had the hearts lightened and
desseosos para hazer la obra de dios
eager to do the work of god
“they announced through the shops that they should get armed as bets as they could and came to fight around the city very strongly, that they had their hearts in flames and eager to do the work of god” (13th c., VLT)

Stative possessive examples with *tener* + PTCP, which are grammatical in Modern Spanish, are found in (59), (60) and (61). The stative possessive interpretation is clear in example (59), where the locative modifier *en tierra* ‘in the ground’ can only be interpreted as the place where the entity (the kneels) is staying.Likewise, in example (60), adverbial *bien* ‘well’ points to the adjectival character of the participle, characteristic of stative constructions.

(59) *el sancto Obispo fizo su oracion al nuestro sennor & dixo*
the saint bishop made his prayer to the our lord and said
teniendo *los inoios* ***fincados*** *en tierra*
having.POSS the knees kneeled in ground
“The saint Bishop prayed to the Lord and said having his knees on the ground” (13th c., EE1)

(60) *la primera que fagan buen gesto quando fablaren en* ***tener***
the first that make good gesture when talked in have.POSS
su cuerpo bien ***asosegado*** *egual mente de gujsa que non*
their body well calmed down likewise of manner that not
apremje mucho las sobre çejas
press much the over eyebrows
“The first is that they make a good gesture when talking, in that they have their body calmed down in a way that it is not pressing too much over the eyebrows” (13th c., CC1)

(61) *los moros* ***tenian*** *sus galeas* ***cubiertas*** *de mantas de lana*
the moors had.POSS their boats covered of blankets of wood

mulladas en agua et la proas delas galeas enel agua
wet in water and the bow of the boats in the water
"The moors had their boats covered with wooden blankets, which were wet in water; and they had the bow of the boats in the water" (13th c. GC3)

An interesting case is the one illustrated in (62), where despite the fact that the possessive reading of *haber* + PTCP is clear, yet the construction has an experiencer interpretation, similar to the one for the English language illustrated in (54). For example, in (62), *enel mercado de Roma* 'in the Roman market' could be interpreted as the place where the heads of a thousand men were cut (experiencer). A perfect interpretation is not possible, as the subject (*mil hombres* 'a thousand men') cannot be interpreted as the agent of the eventuality denoted by the predicate (*cortadas* 'cut'). Somebody else cut their heads and they suffered the consequences of this eventuality.

(62) *occupo la Çiudat por espaçio de .x. anyos. mas ala fin toda*
besieged the city for space of x years but to the end all
aquella legion enla qual hauia .iiijo. mil hombres todos
that region in the wich had iiijo thousand men all
huuieron cortadas *las cabeças enel mercado de Roma.*
had cut the heads in the market of Rome
"He besieged the city during 10 years; but in the end all that region in which there were 40 thousand men, all these men had their heads cut in the Roman market" (13th c., GCI)

3.4.2 Ambiguous contexts

There are some contexts where there seems to be an ambiguity between a stative possessive and a perfect reading or between causative and a perfect reading.

It is very common to find uses of the periphrasis with *tener* and *haber* in Old Spanish, where both a stative possessive and perfect reading are possible. This is illustrated in (64), for *tener* + PTCP, and in (65) and (66) for *haber* + PTCP. In these examples, entity coded by the subject of *haber* could also be interpreted as the agent of the eventuality described by the base predicate. Similar uses to those in the examples expressed by *tener* + PTCP are perfectly grammatical in Modern Spanish as well; however, this ambiguous interpretation is not available to *haber* + PTCP in Modern Spanish, which it is invariably interpreted as a perfect.

(63) *Mars ha semeiança de omne mançebo. & uestido de armaduras*
Mars has resemblance of man young and dressed of armament

*antiguas de grecia todas uermeias & **tiene** en la mano derecha*
old of Greece all red and has.POSS in the hand right
*una espada **sacada** de la uayna .*
a sword taken of the sheath
"Mars looks like a young man, dressed in old army clothes of Greek, all of them red, and has in his right hand his sword taken from the sheath" (13th c., ACE)

(64) *E mato a Metuo ponposiano. porque dizien las gentes que*
and killed to Metuo Ponposiano because said the people that
***tenie** **escripto** el linage de los emperadores. & **tenie** el mundo*
had.POSS written the linage of the emperors and had the world
***figurado** en pargamino.*
figured in parchment
"And he killed Metuo Ponposiano because people said that he had the linage of the emperors written and the world figured out in parchment" (13th c., EE1)

(65) *Sennor dixo. feziste me: merçed e caridat **As** me oy **sacado**: de*
sir said made me mercy and charity have me today taken of
muy grant pobredat
very great poverty
"Sir, he said, you made me with mercy and charity. You have me today taken from great poverty" (13th c., BER)

(66) *Et otro dia mannana ante que saliessen de casa calçaron*
and other day tomorrow before that wen of house put on shoes
*se los Romeros & querellaron se queles **auyan** sus dineros*
REFL the pilgrims and complained REFL that them had their money
***furtados**.*
stolen
"And another day in the morning before they left home the pilgrims put on their shoes and complained that they had their mone stolen" (14th c., BUR)

Interestingly, other cases with *tener* + PTCP and *haber* + PTCP are ambiguous between a causative and a perfect interpretation. In causative readings the subject of *haber* gets the situation expressed in the complement to come about. From the aspectual point of view, base predicates in these cases are accomplishments, as only this semantic class contains CAUSE in its semantic composition (Dowty, 1979). Examples illustrating these ambiguous uses are in (67) and (68).

(67) *el emperador ualent atorgoles las tierras delas traçias do*
the emperor brave gave them the land of the Tracia where
habitassen non prouidiendo ni suspechando se res del
live not foreseeing nor suspecting REFL nothing of the
tiempo esdeuenidero porque cuydaua se siempre ***auer*** *exerçitu*
time to come because cared REFL always have army
apparellado *contra los barbaros*
prepared against the barbarians
"The brave emperor gave the land of Tracia to them, where they could live not beeing afraid of the future, as he would always take care of having an army prepared to fight against the barbarians" (13th c., GCI)

(68) *Capitulo .xxij. de commo el Rey don alfonso querjendo partir para*
chapter xxij of how the king mr Alfonso wanting leave to
el jnperio ***ouo*** *cartas* ***escriptas*** *en araujgo las quales son aqui*
the empire had letters written in arab the which are here
declaradas.
declared
"Chapter 32 On how the King Mr Alfonso, wanting to leave for the empire, had the letters written in Arab, as they are declared here" (13th c., CAX)

3.5 Other contexts of use of *estar*, *haber*, *ser* and *tener* in Old Spanish

So far it has been shown that in Old Spanish periphrases with *ser*, *estar*, *tener* and *haber* held verbal and adjectival passive interpretations, as well as perfects and stative possessive readings. However, apart from these uses in participial constructions, *ser*, *estar*, *tener* and *haber* exhibited other uses in Old Spanish. These include contexts of use with locative prepositional phrases, gerunds, and with individual-level (IL) and stage-level (SL) predicates, and propositions.

3.5.1 With IL or SL adjectives and NPs

As it is well-known, in Modern Spanish *ser* combines with IL predicates whereas *estar* appears with SL predicates (Escandell-Vidal and Leonetti, 2002; Marín, 2000, 2004, 2010). In Old Spanish, in contrast, both copulas could appear with IL and SL predicates.

For example, both *estar* and *ser* are found with adjectives and nouns describing an occupation, whereas in Modern Spanish this type of nouns and adjectives

can only combine with *ser*. See some examples of IL adjectives and nouns with *estar*, which are not grammatical in Modern Spanish, are in (69) and (70).

(69) (*Estar* with IL adjectives)

a. *Hy* ***estauan*** ***contrarios****; los tiempos por yguales.*
There were opposites the times as equals
"There were the opposite the times as equals" (13th c., ALX)

b. *E pareçe que las dichas tasas* ***estan*** ***rrazonables***
and seems that the above mentioned taxes are.LOC reasonable
E ordenamos & mandamos que aquellas se guarden
and ordered and commanded that they REFL keep
"And it seems that the above mentioned taxes are reasonable and thus we ordered that they are kept" (15th c., MTV)

(70) (*Estar* with IL nouns; 13th c. FAR, IJ8)

a. *qui hereda los bienes del padre. o del parient qui aura*
who heirs the good of the father or of the relative who have
estado ***ladron***.
been.LOC thief
"The one who takes over the possessions of the father or the relative will be a thief" (13th c., FAR)

b. *fue el primer sacerdot [...] de babilonia* ***estando*** ***sabidor*** *&*
was the first priest ... of Babilonia being.LOC expert and
maestro *dela ley de moysen*
master of the law of Moysen
"He was the first priest of Babilone to be an expert and master in the law of Moyses" (13th c., IJ8)

Conversely, *ser* in Old Spanish combines with SL predicates. See some examples in (71). In Modern Spanish, only *estar* can appear in these contexts.

(71) (*Ser* with SL adjectives)

a. *Vieno a ueer ector; se* ***era*** ***biuo*** *o muerto. Fallo lalma yda;*
came to see Ector REFL was alive or death found the soul gone
& finado el cuerpo.
and dead the body
"Ector came to see whether he was alive or death. He found the soul gone and the dead body" (13th c., ALX)

b. *De como deuen fazer fuera de la tinaia un logar en que quepa*
of how must make out of the jar a place in which fits

agua poca. & que ***sea*** *toda uia* ***lleno*** *de agua.*
water a bit and that be still full of water
"On how to make out of a jar a place where a bit of water can be kept, and still be full of water" (13th c., AST)

Similar examples to those we find in Modern Spanish are in (72), for *ser*, and in (73), for *estar*.

(72) (*Ser* with IL predicates)

a. *Ca desde que fue Rey non auie guerreado. Sy estonçes fues*
because since that was King not had fought if then was
morto. fuera bien auenturado El quanto ***era*** ***rico****. tant era*
killed was well lucky he how much was rich so was
poderoso.
powerful
"Because since he had become a King he had not been fighting. If he were killed then, he would have been lucky. He was as rich as powerful" (13th c., ALX)

b. *Platon el mejor philosopho & mas sabio de aquel tiempo:*
Platon the best philosopher and more wise of that time
era ***maestro*** *de nicophoris fijo de aquel Rey.*
was master of Nicophoris son of that King
"Plato was the best and wiser philosopher of that time; he was master of Nicophoris, son of that King" (13th c., BDS)

(73) (*Estar* with SL predicates)

a. *E los que aqui vos acahescistes de* ***estar*** ***presentes*** *sabed que*
and the that here you happened of be.LOC present know that
el saber que es grand don de aquel que nos fizo.
the knowledge that is great gift of that that us made
"And those of you who were present should know that knowledge is a great gift of the one who made us" (13th c., BDS)

b. *Ca el vn dia* ***estara*** ***vazio*** *de gente & otro* ***estara***
because the one day be.LOC empty of people and other be.LOC
muy ***lleno***
very full
"because one day it will be void of people and other it will be very full" (13th c., CD1)

On the other hand, in Old Spanish, both *haber* and *tener* can appear with SL-adjectives. See some examples with *haber* in (74) and with *tener* in (75). In

these contexts *haber* and *tener* describe some state of possession, in a similar way as stative possessive uses with participles, as described in the previous section. Modern Spanish *haber* can never appear in this context.

(74) (*Haber* with SL adjectives)

a. *E la zaraffa es bestia grande fecha como cierua. & **ha** el*
and the giraffe is beast big made as deer and has the
*pezcueco muy **luengo** & la cabeça chica & los oios muy*
neck very long and the head small and the eyes very
fremosos & las piernas delante muy luengas
beautiful and the legs front very long
"And the giraffe is a beast which is made as deer. It has the neck very long and the head small and the eyes beautiful and the front legs very long" (13th c., ACE)

b. *don nuño fablo con estos mandaderos si podria **auer***
mr Nuño talked with those servants whether could.3SG have
*aquellos caualleros **sueltos***
those knights released
"Mr Nuño asked to those servants whether he could have those knights released" (13th c., CAX)

(75) (*Tener* with SL adjectives)

a. *Non te atreuas en cosa en que non **tengas** los pies **firmes***
not you dare in thing in that not have.POSS the feet stable
"You should not dare to put your feet over something where your feet are not stable" (13th c., CD1)

b. *mando que la [casa] cerrassen & la **touiessen** muy **limpia***
ordered that the house closed and her had.POSS very clean
& muy guardada.
and very kept
"He ordered that the house should be closed and that they have it very clean and well kept" (13th c., VLT)

3.5.2 With locative prepositional phrases

Since the earliest centuries both *estar* and *ser* could appear in contexts with oblique prepositional phrases describing the location or place of the entity expressed by the grammatical subject. This is illustrated in sentences in (77) and (76). In Modern Spanish *ser* can only very rarely appear with spatio-temporal modifiers.

(76) (*Ser* with locative prepositional phrases)

a. *Commo lo dixo el çid assi lo han acabado Mynaya albarfanez*
as it said the Cid so it have finished Mynaya Albarfanez
fuera ***era*** ***enel campo*** *Con todas estas yentes escriuiendo &*
away was in the field with all these people writing and
contando
telling
"As the Çid said, so they finished it. Mynaya Albarfanez was out in the field with all these people writing and telling" (12th c., CID)

b. *Et dezimos que la linna que* ***es*** ***en la tierra*** *so el*
and say that the line that is in the ground below the
cerco dell yguador del dia; es la linna a que dizen linna
mid-point of the day is the line to that say line of the
de la egualdad.
equality
"And we say that the line which is on the ground below the mid-point of the day is the line which is called the line of equality" (13th c., ALB)

(77) (*Estar* with locative prepositional phrases)

a. ***Estando*** ***en la cruz*** *vertud fezist muy grant*
being.LOC in the cross virtue made very big
"You made a great virtue by being in the cross" (12th c., CID)

b. *Quando ell alffil* ***esta*** ***en el tablero***. *si algun peon esta*
when the bishop is.LOC in the table if some pawn is.LOC
depos ell a una casa
after it to a house...
"When the bishop is in the table, if some pawn is after it at a house..." (13th c., ACE)

As illustrated in (78), in Old Spanish *ser* could also combine with prepositional phrases describing the location of the entity in a more abstract state of affairs or state of mind can also be found. These cases are also not possible in Modern Spanish, when only *estar* can appear in these contexts.

(78) (*Ser* with abstract locative prepositional phrases)

a. *otrossi los omnes que son uieios & flacos. o los que han*
then the men that are old and skin or the that have
sabor de auer sus plazeres apartadamientre porque non
pleasure of have their pleasures away because not
reciban en ellos enoio nin pesar; o los que ***son*** ***en poder ageno***
receive in them anger nor annoy or the that are in power strange

assi como en prision
so as in prison
"And then the men who are old and and skinny or those who like enjoying their pleasures away, so that they do not suffer anger nor pain, or those who are in foreign power or in prison" (13th c., ACE)

b. *E quanta mengua* ***es en la oraçion*** *de aquellos que sin*
and many reduction is in the prayer of those that without
amor de Dios & de las almas del cuerpo de Ihesu Cristo fazen
love of God and of the souls of the body of Jesus Christ make
oraçion a Dios Padre por ellas.
prayer to God Fater for them
"And there is reduction in those who pray without without loving God" (14th c., CF3)

c. *Por la qual cosa los xristianos xvi ccxxxviii* ***fueron en grant***
for the which thing the christians xvii ccxxxviii were in great
pensamiento *et* ***en grant paor***
thought and in great fear
"The Christians were in great worry and fear because of that" (14th c., CQ2)

Similar examples with *estar*, which are grammatical in Modern Spanish, are found in (79).

(79) (*Estar* with abstract locative prepositional phrases)

a. *por dios e caridat* ***En oracion estaua***
for god and charity in prayer be.LOC
"He was praying for God and charity." (13th c., BER)

b. *E entro a la camara do* ***estava*** ***en este pensamiento***
and entered to the chamber where be.LOC in this thought
"And he came into the chamber where he was thinking." (13th c., CZP)

c. *Ellos* ***en sto estando*** *don auien grant pesar.*
they in this being.LOC where had great sorrow
"They were in that, where they had great sorrow" (12th c., CID)

3.5.3 With gerunds

Both *estar* and *ser* can combine with gerunds holding a progressive interpretation. See examples with *ser* in (80) and (81), and with *estar* in (82) and (83). This contrasts with Modern Spanish progressive, which can only be expressed with

estar.

(80) *E otro dia quando sopieron que* ***<u>era</u>*** ***entrando*** *en cordoua*
and other day when knew that was entering in Cordoba
vinjeronse el rrey don alfonso & el Rey abenyuçaf con sus
cameREFL the king Mr Alfonso and the King Abenyuçaf with their
huestes a çercar la villa
armies to fence the village
"And the following day, when they knew that he was entering in Córdoba, the King Don Alfonso and Abenyuçaf came with their armies to fence the village." (13th c., CAX)

(81) *E despues ponlos en su vara E* ***<u>sea</u>*** ***Ardiendo*** *candela toda la*
and then put them in their stick and is burning candle all the
noche
night
"And then, put them in their stick and let the candle be burning during all night long." (14th c., CET)

(82) *E tornaron se al palaçio pora la cort Myo çid por sos*
and went back REFL to the palace for the court Mio Cid for his
yernos demando & no los fallo Maguer los ***<u>estan</u>***
brothers in law asked and not them found although them stay
lamando *ninguno non responde.*
calling nobody not responds
"And they went back to the palace; Mio Cid asked about his brothers in law but he could not find them. Although they are calling them, nobody responds." (12th c. CID)

(83) *De la figura de Mercurio. Mercurio ha semeiança domne mançebo*
of the figure of Mercury Mercury has resemblance of man young
uestido de pannos de muchas colores. & que ***<u>esta</u>*** ***escriuiendo*** *en*
dressed of clothes of many colours and that is.LOC writing in
un Libro.
a book
"About the figure of Mercury. Mercury looks like a young man wearing colourful clothes and who is writing in a book." (13th c., ACE)

(84) ***<u>ouo</u>*** *muy gran pesar* ***veyendo*** *que esto uenja por mengua della enbio*
had very great sorrow seeing that this came for drop of her sent
sus mandaderos al Rey de noruega
his messengers to the King of Norway
"He had great sorrow seeing that this came as she fell and he sent his

messengers to the King of Norway." (13th c., CAX)

3.5.4 With propositional clauses

Lastly, it should be mentioned that since the earliest centuries the type of internal argument, coded as the object, of *haber* and *tener*, could also be a proposition. See examples with *haber* in (85) and (86) and with *tener* in (87) and (88). In Modern Spanish only *tener* can be used in these contexts, and even now these uses are not very productive.

(85) *Los omnes que se uezan tal uida mantener. Son malos ganadores;*
the men that REFL use so life keep are bad winners
no ***an*** ***onde lo auer.***
not have where it have
"Men who are used to live that type of life are bad winners; they do not have where to have." (13th c., ALX)

(86) *la entregaria aben alhamar que gela diera por que se non*
her would-deliver Aben Alhamar that REFL.her gave because REFL not
perdiese en su tenençia de abenyuçaf njn los moros ***oujesen***
lose in his ownership of Abenyuçaf nor the moors had
quele dezir
that him say
"Aben Alhamar would deliver it to them. He would give it to them because he did not want to loose Abenyuçaf and also he did not want the moors had anything to tell him." (13th c., CAX)

(87) *conel daño que les han fecho los que estan en granada*
with the pain that them have done the that are.loc in Granada
tengo que muy poco duraria la guerra *mas non me semeja que*
have.poss that very little last the war but not me seems that
ay njnguno que vos diga esto
there is nobody that you tell this
"With the pain that those who are in Granada have made to them, I think that the war would last for a very few time, but there seems to be the case that nobody is telling you this" (13th c., CAX)

(88) *Paros me delantre; un ombre reuestido. En que ombre lo llamo;*
stopped me in front a man dressed in that man it called
tiengo me por fallido. ***Tiengo que era angel; del çielo***
have.poss me for failed have.poss that was angel of the sky

deçendido.
descended
"A dressed-up man stopped in front of me. I do not know who called him. I think it was an angel fallen from the sky" (13th c., ALX)

3.6 Interim summary

To sum up, the range of interpretations of participial constructions in Old Spanish, as just presented, is summarised in Table 3.1. As can be seen from this table, the perfect interpretation could be instantiated by all four periphrases. The other interpretations, in contrast, were realized by two participial constructions. Specifically, the stative possessive interpretation was expressed by *haber* + PTCP and *tener* + PTCP, and verbal and adjectival passives were instantiated by *estar* + PTCP or *ser* + PTCP.

	Perfect	StaPoss	VerPass	StaPass
Haber + PTCP	+	+		
Estar + PTCP	+		+	+
Tener + PTCP	+	+		
Ser + PTCP	+		+	+

Table 3.1: Interpretations of participial constructions in Old Spanish. In this table, the label *StaPoss* stands for 'stative possessive', *VerPass* for 'verbal passive' and *StaPass* for 'adjectival passive' or 'stative passive'.

In contrast, in Modern Spanish each auxiliary is clearly associated with only one interpretation. Specifically, *haber* + PTCP invariably expresses the perfect and *tener* + PTCP holds a stative possessive interpretation, whereas *ser* + PTCP and *estar* + PTCP instantiate verbal and adjectival passives, respectively. The causative and experiencer uses to *haber* and *tener* have been lost. This is summarised in Table 3.2.

	Perfect	StaPoss	VerPass	StaPass
Haber + PTCP	+			
Estar + PTCP				+
Tener + PTCP		+		
Ser + PTCP			+	

Table 3.2: Interpretations of participial constructions in Modern Spanish.

On the other hand, apart from passive, perfect and stative interpretations of participial constructions, in Old Spanish there were other uses to *haber*, *ser*, *estar* and *tener*. These uses covered contexts with locative prepositional phrases, gerunds, propositional clauses and with both IL and SL predicates.

It is relevant at this point to note that in this area the distribution of the interpretations has substantially changed as well. Specifically, in contrast with Old Spanish, in Modern Spanish *haber* can never appear with SL or IL predicates, locatives, or gerunds and propositional clauses, which are in turn only possible with *tener*. As for copulas, both *ser* and *estar* show some of these uses, but in complementary distribution: whilst *ser* can only appear with IL predicates (and only very few locative phrases), *estar* combines with SL predicates, with gerunds to form progressives and with locative prepositional phrases.

Some of the questions that may now arise are as follows: How could these periphrases, and more generally *tener*, *estar*, *ser* and *haber*, instantiate so many different interpretations in Old Spanish? How and why did their interpretations change over time? In the last sections of this chapter I try to begin to answer these questions.

3.7 Semantic underspecification and the interpretations of *haber*, *tener*, *estar* and *ser* in Old Spanish

Intuitively, the most logical possibility to explain the whole range of contexts that *haber*, *tener*, *ser*, *estar*, could participate in, giving rise to a wide range of different interpretations, is that in Old Spanish these predicates had an underspecified semantics. On this view, each predicate would have disjunctive or incomplete semantic features, but in the course of putting together a syntactic representation a particular option was selected, possibly via interactions with context.

The meaning of *haber* and *tener* would be that of a highly underspecified two-place relation between an entity and either an entity or a situation. In Old Spanish, both *haber* and *tener* would be relational elements very much like non-auxiliary uses of English *have* (Cowper, 1989; Guéron, 1986; Harley, 1998, among others). Likewise, the semantics of both *ser* and *estar* in Old Spanish would that of one which denotes an identity function over properties, similarly to Modern English *be*, as proposed by (Cann, 2006).

3.8 Complex variation and regularization

So far we have seen that in Old Spanish there were four different interpretations which could be assigned to four forms or periphrases almost interchangeably. On

the one hand, we saw that in Old Spanish there was great variation in the perfect. The perfect could be morphosyntactically instantiated by periphrases formed with *tener*, *haber*, *estar* and *ser*. All these periphrases appeared in similar contexts, specially regarding adverbial modification, and combinations with some classes of verbs. Such variation in the perfect does not longer exists in Modern Spanish, which only expresses the perfect with *haber*. On the other, in Old Spanish there was variation in passives as well, in that both verbal and adjectival passives could be expressed by *ser* and *estar*. Lastly, there were other interpretations to periphrases formed with *haber* and *tener* (stative possessive, causative, experiencer) which stem from the possessive meaning of these verbs in the earliest centuries.

This would result in a **multiple label system** which is starkly different from Modern Spanish, where each each periphrasis unambiguously realizes one interpretation. The great linguistic variation in the expression of these interpretations, as just described, was complex and bidirectional, in the sense that there was variability not only in the morphological instantiation for these interpretations, as each interpretation could be expressed by two or more periphrases, but also in the number of interpretations that each participial construction could express, ranging from two to three. Furthermore, this variation extended to uses of auxiliaries outside participial constructions. Informally, from the semantic point of view, Old Spanish participial constructions behaved, on the one hand, like synonyms appearing in similar contexts and exhibiting the same interpretations and, on the other, like highly polysemic forms in that each form could hold different meanings or interpretations.

In this respect, the system in Old Spanish exhibited highly **complex or unpredictable variability** in the perfect, passive and stative marking: four possible markers were used interchangeably to express the same grammatical interpretations. This complexity grows bigger if one considers ambiguous occurrences in which one cannot decide on the basis of context the interpretation that should assigned to the construction, as in (89), where both a perfect and a stative possessive interpretation are possible. In the perfect interpretation (i) the sentence describes the eventuality of noting something down in a book, in (ii) the sentence describes the stative situation of having a book where something is noted down.

(89) *Et assi lo **auemos** nos **puesto** en este libro.*
and so it have we put in this book
(i) "And we have put it like this in this book."
(ii) "We have it put like this in this book." (13th c., ALB)

Crucially, the kind of variation observed in Old Spanish in the expression of perfects, passives and statives does not exist in any modern language variety with participial constructions. Even in languages like French, in which one auxiliary

(*avoir* 'have') can express both perfects and stative possessives, and other auxiliary (*être* 'be') realizes both verbal and adjectival passives and even the perfect, there is clearly variation, but this is largely predictable. For example, in the case of the French perfect, most unacussatives select *être* whereas transitives and unergatives select *avoir*. Furthermore, in these languages there are only two forms at most exhibiting these interpretations, whereas in Old Spanish there are four morphological instantiations for the same interpretation. Likewise, the complexity that we observe in Old Spanish does not exist in Modern Spanish, where each periphrasis realizes one and only interpretation; namely *haber* + PTCP realizes the perfect, *ser* + PTCP the verbal passive, *estar* + PTCP the adjectival passive and *tener* + PTCP is the expression for stative possessives.

From the language change point of view, a system with multiple morphological instantiations for perfects, passives and statives has developed into another with unique forms for such interpretations. From this system, synonymy and multiple labels or forms have disappeared. The complex variability has been eliminated from language by specializing the meaning of each participial construction: over time, each interpretation comes to be associated with a particular periphrasis.

In uses other than participial constructions, auxiliaries in Old Spanish also showed a considerable amount of complex variation. Both *ser* and *estar* could combine with IL and SL predicates and with locatives and gerunds to form progressives; and both *tener* and *haber* could combine with SL predicates and propositions. Also in this area, complex variability has been eliminated to a certain degree over time, as *ser* eventually came to be combined only with IL predicates, whereas *estar* was specialized to combine with SL predicates, locatives and gerunds, and *tener* appears only with SL predicates in Modern Spanish.

On the basis of these findings, a type of change that could explain this would be one in which there has been some kind of **regularization** leading to the specialization of each participial construction to express one and only one interpretation. The end result of such process of change is a linguistic system which still exhibitis variability, but that variability is unequivocal: the choice of participial construction is conditioned on the linguistic distribution, such as the type of verb and other contextual characteristics. Thus, the outcome of this regularization change will be different for each auxiliary:

- *haber* becomes the perfect auxiliary providing only grammatical information about person, number and tense and can, thus, only appear in combination with an the past participle, which provides the thematic and lexical information to the construction.

- *tener* remains as a stative possessive, combining with SL predicates.

- *estar* becomes the copula used for SL predicates, locative prepositional

phrases, as well as adjectival passives, and progressives.

- *ser* becomes the copula used for IL predication and verbal passives.

Such regularization change is expected under the well-known property of natural languages that variation tends to be predictable. In general, in systems no two linguistic forms will tend to occur in precisely the same environments and perform precisely the same functions (Givón, 1985). From the language change point of view, a direct consequence of this claim is that given a certain amount of such complex variation, consistency may eventually arise in language as a result of some regularization process. From the language change point of view, this type of change conforms to the **One-Meaning-One-Form Principle**, also called **isomorphy** or Humboldt's universal: maximally efficient system avoids polysemy, or two forms with similar meanings and different forms, and homophony, two unrelated meanings getting the same form (Anttila, 1972).

Similar processes of regularization have been attested for older language varieties other than Old Spanish. In particular, in grammaticalization phenomena besides changes in the semantic content of forms (semantic bleaching) there is often the case that one morphological expression prevails to express one interpretation, whereas other competing forms which in previous stages of the language described similar meanings eventually disappear (Hopper and Traugott, 2003). In this sense, specialization has been considered as the process of reducing the variety of formal choices available as the meanings assume greater generality.[1]

For example, specialization may account for change in Indo-Aryan compound verbs, where the range of so-called **vector verbs** was eventually reduced as the meaning of these verbs generalized, with the consequent loss of some of these verb forms (Hook, 1991). Another well-known example of specialization in grammaticalization phenomena is the development of negation in French. In contrast with Modern French, where negation is expressed by *ne...pas*, in Old French a variety of nouns describing some quantity, e.g. *point* 'dot, point', *mie* 'crumb', *pas* 'step', *gote* 'drop', among others, could be placed after *ne* plus verb in order to reinforce negation. By the 17th century, only *pas* (and in less contexts, *point*) becomes fully grammaticalized out of the old array of forms which were available in the earliest centuries.

Theories of syntactic change also acknowledge the vast periods of variation that there are between two word orders (Kroch, 1989b, 1994, 2005). This type of optionality tends to disappear in languages as well. For example, Late Middle

[1] The term **specialization** has also been used, in a different sense to the one used here, to refer to the traditional type of semantic change which involves the restriction of a word's semantic field, or narrowing in the number of referents included by the term, e.g. the change from Old English *meat* 'food' to Modern English *meat* 'flesh of an animal as food'.

English manifests a variation between verb-second (V2) and simple SVO word order that it is not found elsewhere among V2 languages (Kroch, 1989a). Similarly, Ancient Greek evolves from SOV to SVO word order in the centuries between the Homeric period and the New Testament, with extensive period of variation between the two word orders (Taylor, 1994).

It has been proposed that such periods of transition in syntactic change proceed via **competition** between grammatically incompatible options which substitute for one another in usage, and that this behavior shows that syntactic heads behave like morphological doublets and are therefore subject to the **blocking effect** (Aronoff, 1976), which excludes doublets and more generally any coexisting formatives that are not functionally differentiated. Thus, such variation between mutually exclusive grammatical options reflect competition between grammars.

Going back to our case study, the variation in the selection of the different auxiliaries in participial constructions to express a wide range of interpretations in Old Spanish suggests that at this stage participial constructions system was a system in transition. In this case study, competition takes place, on the one hand, between participial constructions to be the morphosyntactic instantiation of perfects and, on the other, to instantiate stative possessive and both verbal and stative passive interpretations. As a result of a regularization change, and possibly motivated by this competition to express the same interpretations, each participial constructions specializes to express one and only one interpretation.

Crucially, this variation is not exclusive to old language varieties. A similar type of variation has been found at earlier stages of language acquisition (Birdsong, 1999; Johnson *et al.*, 1996; Hudson Kam and Newport, 2005). Specifically, this complexity has been observed in creole or pidgin languages and in other emerging languages. These languages contain variability, particularly in the linguistic expression of grammatical categories, which is not typical of natively acquired languages. For example, speakers are inconsistent in their use of morphological markers or word order. In later stages of emerging languages, in general, it is also often the case that consistent variation is found precisely in the same contexts where before there had been unpredictable variation. For example, during the acquisition of sign languages from teachers using inconsistent grammatical forms of sign it has been attested that learners tend to regularize (Singleton and Newport, 2004).

This idea that similar types of changes occur both in language acquisition processes and language change is not new. It has been largely suggested that the same mechanisms may underlie diachronic language development (Hudson Kam and Newport, 2005), and even that the directionality that has been found in so many language developments can be explained by more general mechanims of language acquisition such as priming or alignment (Jäger and Rosenbach, 2008).

The question now arises as to how specifically this regularization change in

participial constructions take place over time, from Old to Modern Spanish. Does this regularization change happens abruptly or gradually? If change in the interpretations of participial constructions happens as a result of regularization, then we expect to find a pattern for changes in these forms over time alike to those observed in language acquisition studies for similar phenomena.

For example, Smith and Wonnacott (2010) study the elimination of complex variation through generations of learners in an artificial language set-up. Specifically, they demonstrate that **iterated learning** (where learners observe and learn a linguistic behaviour which is itself learned) as it happens through **diffusion chains** (where the output of one learner forms the input to the next learner in a chain of transmission) can produce linguistically-conditioned stable variability. More specifically, they explore how the variability in choice of plural markers becomes predictable through chains of speakers, with usage becoming conditioned on the noun being marked, even though such regularization change was not observable in the first chains of speakers.

The aim of the following chapter, will be precisely to explore change in participial constructions as it happens over time and to check how changes in the frequency of these periphrases conform to regularization process as previously observed in language learning studies.

3.9 Summary & Concluding remarks

In this chapter I presented a in-depth (qualitative) analysis of participial constructions in the earliest centuries as well as a tentative general proposal about the regularization type of change which could give rise to the interpretations of these periphrases in Modern Spanish.

In the following chapters I will ellaborate this proposal further and test it on the corpus using quantitative data. In this respect, I will be also exploring how change happened throughout the centuries and begin to look for answers to the question as how changes can spread over time through generations of speakers.

4

FREQUENCY CHANGES

4.1 Introduction

The goal of this chapter is to report the results of a quantitative analysis of change in participial constructions, based on frequency data of these periphrases retrieved from the corpus. The quantitative study reported here addresses the basic question of how the regularization change in participial constructions, as proposed in the previous chapter, takes place over time. Specifically, I investigate the question of how regularization is reflected in the frequency changes of participial constructions. Based on previous probabilistic approaches to language learning and development of inconsistent variation in grammar (Reali and Griffiths, 2009; Smith and Wonnacott, 2010), I explore whether changes in the usage frequency of participial constructions fit expectations from regularization processes.

In historical linguistics frequencies have for a long time been reported to indicate changes in the usage of the given linguistic expression across time (Ellegard, 1953). There are, however, different views on what changing frequencies actually indicate.

Studies in the functional framework consider frequency to be a mechanism for change, and also *prima facie* evidence of the degree of grammaticalization of a given form (Heine *et al.*, 1991; Bybee *et al.*, 1994; Bybee, 2003, 2010; Hopper and Traugott, 2003; Fischer, 2010, among others). For example, Bybee (2003), who deals with frequency changes in grammaticalization phenomena, claims that the dramatic increase in frequency so often encountered in this type of change is both a result and a primary contributor or active force in instigating grammaticalization changes. Building on Haiman (1994) and Boyland (1996), she claims that frequency and repetition lead to the automatization and semantic bleaching of the given form.

In contrast, in the generative tradition, frequency changes are considered a

reflection that some change has occurred in the grammar of the language (Kroch, 1989b, and subsequent work). According to Kroch and other studies following his general approach, variation between mutually exclusive grammatical options reflects competition between grammars. Changing rates in usage frequency reflect such competition and the gradual replacement of one of the options by the other, which stands up as the eventual winner.[1]

An alternative view on the role of frequency in language comes from the field of language acquisition. Recently, there has been found evidence that statistical information about frequency distributions of words and expressions play an essential role in the learning and development of a language (Gopnik *et al.*, 2004; Xu, 2007, 2011, to name just a few). The key idea in these studies is that learning mechanisms are a set of rational, inferential, and statistical mechanisms. This new perspective blends elements of a constructivist account of development with the account of learning as rational statistical inference underlying probabilistic models of cognition. Thus it provides a way to integrate symbolic representations with statistical learning, combining research from two related areas from psychology and cognitive science: on the one hand, probabilistic models of cognition that provide new insight into the nature of human learning and inferencing (Chater *et al.*, 2006; Oaksford and Chater, 2007, among others); on the other, developmental psychologists that go beyond the severe fracture underlying the nativist and empiricist debate on language development. For a survey see the editorial and papers included in the special issue at *Cognition* (Xu and Griffiths, 2011). Probabilistic models of language acquisition have also explored the way language learning happens through generations of speakers. Very informally, in these models language development proceeds in the following way: the learner first observes the expression in the language input, makes an inference about its probabilities, and then produce forms for the same interpretations with a certain frequency, which are then observed by the next learner.

The link between language acquisition and language change has long been recognized, and therefore it seems a reasonable strategy to build on advances made in this field to explore language development as it takes place over longer periods of time. However, to my knowledge, so far historical linguistics research on language change has not incorporated yet knowledge coming from these studies. A programmatic study in this direction in the field of historical linguistics is proposed by Rosenbach and Jäger (2008), who suggest that general cognitive mechanisms such as priming can explain particular changes, as for example the unidirectionality observed in grammaticalization. In this respect, with the case study presented here I hope to contribute to begin to integrate these methods in

[1] More standard generative approaches do not share the view that variation must be accounted for by positing competition between two grammars.

the field of historical linguistics.

The ordering of contents in this chapter is as follows. In Section §4.2, I summarize the main results obtained in probabilistic studies about language learning and development of complex variation. In Section §4.3, I first describe the method used to obtain participial constructions from the corpus and present the results of an evaluation over the retrieved data (4.3.1), and then I present the statistical method used to test the results for significance (4.3.2). Lastly, in Section §4.4 I report the results of the quantitative study on changes in the frequency of participial constructions.I conclude this chapter in Section §4.5 with a summary and some concluding remarks.

4.2 Frequency and regularization

When learners are faced with complex variation two different patterns have been observed in language acquisition studies: **probability-matching** or **regularization** (Hudson Kam and Newport, 2005). When learners probability-match, the utterances produced by learners are proportional to the frequency of the linguistic variants in the input language, as they are produced by speakers. Conversely, regularization happens when learners collapse complex variation towards a more consistent rule, in which one variant becomes the vast majority in the population, and thus other variants competing for the expression of the same meaning are eliminated. The strategy followed by learners depends on several factors, including the age of learners as well as the degree of complexity of the input language. For example, Hudson Kam and Newport (2005) trained children and adult participants on artificial languages where determiners occurred with nouns with varying probabilities. They found that children regularize the complexity in the input, producing consistent patterns that were not the same as the training stimuli. Conversely, they found that adults produced utterances with probabilities proportional to their frequency in the training (probability matching).

On the other hand, it has been observed that there is some bias towards regularization when the acquisition of complex variation is observed through generations of speakers (Reali and Griffiths, 2009; Smith and Wonnacott, 2010). In the long term, it has been observed that learners tend to regularize, even if such tendency is not so easily tracked in one generation of speakers, which may mostly probability-match (Reali and Griffiths, 2009; Smith and Wonnacott, 2010).

On the basis of recent studies on regularization in language acquisition as it happens through generations of speakers, it is also possible to draw some expectations about how the frequency distributions of particular forms will change over time given a considerable amount of complex or unpredicatible variation.

Changes in the frequencies are an approximate measure for how participial

constructions are changing over time. If we imagine a similar set-up for language development as the ones described before for the studies just summarised, the problem could be framed in the following terms: the frequency with which a participial construction is used at particular point in time is the input language to which the next generation of speakers is exposed. The output produced by this following generation is available to the second generation of speakers, whose output frequencies will be available in later documents, and so on and so forth.

On the basis of the well-known patterns of word frequencies in natural languages, it is possible to propose some hypotheses about the types of changes we could possibly find in such frequencies. As it is well-known (Zipf, 1935, 1949), there are similar patterns in the distribution of word frequencies across languages. Specifically, there is a low number of functional words, e.g. determiners, in language occurring many times, whereas there is a high number of contentful words (such as nouns) in language ocurring fewer times.

Thus, from the language change perspective, it follows that a change in the frequency pattern of a word may indicate a change in its class. In particular, if the relative frequency of appearance of a particular lexical item increases significantly across a wide range of text types over time, this is usually a clear indication that this particular expression is becoming more functional in nature. In other words, it is changing its grammatical status from a lexical category into a functional category. Conversely, if there is a significant decrease in the frequency of a linguistic form, it might be a sign that this form is becoming more "lexical" in nature (which is rather infrequent in processes of language change) or, more likely, that it is losing some of its functional interpretations.

Grammaticalization, however, is not the only possible cause for an increase in frequency. Linguistic expressions may also become more frequent as they are gradually established in the language. Specialization to a specific interpretation could lead either to a decrease in frequency (fewer applicable contexts) or to an increase (as it becomes the default construction for this interpretation and other competing constructions decrease in frequency).

In the case study of participial constructions, thus we expect to find a significant increase in the frequency of *haber* + PTCP, as it is undergoing a reanalysis from a possessive stative lexical verb to an auxiliary expressing a perfect interpretation of the event denoted by the main verb (grammaticalization). Conversely, as *ser* + PTCP eventually loses the perfect and adjectival passive interpretations, we expect to find a significant decrease in its frequency. On the basis of the changes in the interpretations of *estar* + PTCP and *tener* + PTCP observed in the previous chapter, it is not so easy to draw expectations about changes in their frequency distributions. However, probably the fact that they cease to appear in verbal passive and perfect contexts suggests that we should also expect some decrease in their frequency.

But before presenting the results of the study of changes in the frequencies of participial constructions, in the following section I will present the method used to retrieve such frequency data from the corpus and the procedure used to test frequency changes for significance.

4.3 Method

4.3.1 Data

In order to conduct this study, frequency counts of all occurrences of *estar* + PTCP, *haber* + PTCP, *tener* + PTCP and *ser* + PTCP were obtained from the corpus using the *IMS Open CorpusWorkbench*[2] and analyzed with the open-source statistical software R (R Development Core Team, 2010). The CQP macro language in the *IMS Open CorpusWorkbench*, which makes use of regular expression-like syntax and applies automated pattern matching techniques, was used to retrieve the examples.

Being periphastic expressions, automatic retrieval of participial constructions is a challenging task *per se*, especially if one considers the fact that in Spanish the participle can be placed both after or before the auxiliary, and a number of elements can be placed between both elements. The syntactic variability is even bigger in the earliest centuries, as expected also from the fact that these constructions were changing significantly in the Middle Ages.

The queries used to retrieve occurrences of participial constructions have been carefully elaborated on the basis of the observations of auxiliary-participle occurrences in the corpus, paying special attention to the classes of tokens placed between auxiliaries and participles, as well as the categories of tokens to the right and left of these periphrases. For reasons of efficiency in the retrieval process, the same queries have been used to obtain occurrences throughout all centuries and for all 4 participial constructions. A total number of 15 different queries were elaborated to cover the maximum number of cases; 14 of which obtain instances of the periphrases in auxiliary-participle word order, whereas only 1 deals with cases in inverse word order.

Table 4.1 summarises the syntax of these queries. Each column in this table summarizes the main information about each of these queries. The second column indicates the word order of the participial construction instance, and the third column indicates the number of tokens which can be found at most between auxiliary and past participle. The fourth column summarises the classes of the token(s) which are found between auxiliary and past participle, where ';' indicates distinct possibilities and '+' a combination of elements in the given order, with ','

[2] `http://cwb.sourceforge.net/`

separating different alternatives. Lastly, the fifth and sixth columns indicate the contexts restrictions to the right and left of participial constructions, when needed. When 'No' is written, it indicates that the occurrence should not be retrieved if it is found in such context, as it may result in a false case.

Alphabetical characters in this table are abbreviations for the morphological class of the tokens in these queries. The abbreviations used here correspond to the first characters in the part-of-speech (PoS) labels used by the linguistic analyzer (FreeLing), which was used to enrich the corpus with linguistic information, as described in chapter 6. A complete list of the tags can be found in the FreeLing documentation.[3] For example, 'CS' and 'CC' stand for subordinate and coordinate conjunction, respectively, 'N' is noun, 'DP' is possessive determiner, 'DI' is indefinite determiner, 'DA' is definite determiner, 'P' is pronoun and 'PR' stands for relative pronoun, 'Z' is number, 'R' is adverb, 'A' stands for adjective, 'F' is punctuation symbol, and 'S' is preposition. Another abbreviations used here are 'NP', which stands for noun phrase,[4] and 'Aux', which represents *haber, estar, tener* and *ser* in both finite and non-finite verb forms excluding past participles, and 'mente' stands for adverbials in *-mente* such as *generalmente* 'generally'.

As can be seen from this table, in the queries for auxiliary-participle word order up to 19 words of different morphological categories can be placed between, whereas in the query for participle-auxiliary word order only two elements at most can be placed between.

To put an example, query number 12 represents a corpus query to retrieve instances of participial constructions containing from three to seven tokens between both forms, none of which can be verbs, as in (1). Furthermore, this query indicates that there cannot be any personal pronoun, noun, or finite verb form of *ser*, *estar*, *tener*, or *haber* to the right of participial periphrases. This contextual boundary prevents the query from retrieving cases such as the one in (2), where there is no participial construction but the verb *ser* plus a participle (*sennalada*) followed by noun *mientre*, which renders the participle into an adverb, like Modern Spanish *señaladamente*.

(1) *Maestre aristotil; que lo auie criado. Sedia eneste commedio; en su*
master Aristotle who it had raised was in this time in his
camara çarrado. ***Avia*** *un silogismo. de logica* ***formado***.
chamber locked up had a **o**f logic formed
"At that time, master Aristotle, who had raised him, was locked up in his chamber. He had created a syllogism" (13th c., ALX)

[3] `http://nlp.lsi.upc.edu/freeling`

[4] The expression used to represent noun phrases includes the following words: an optional determiner, an optional number of adverbs, an optional adjective, a noun, optional adverbs and optional adjective.

(2) *touieron por meior de començar en las figuras que* ***son*** *a cerca del*
had.POSS for better of start in the figures that are at close of the
zodiaco. & ***sennalada mientre*** *en la figura de caytoz*
zodiac and pointed -ly in the figure of Caytoz
"They thought it was better to begin talking about the figures which are close to the Zodiac, specifically about the figure of Caytoz" (13th c., AST)

The total counts of participial constructions retrieved from the corpus using these queries for each document in the corpus are listed in Appendix B.

In order to evaluate the occurrences obtained using these queries, the number of correctly retrieved participial constructions (precision) was counted. The precision of the queries was obtained over an evaluation of 50 randomnly selected cases of each of the four participial construction both from the 13th and the 20th century, so that the results obtained both in the earliest and latest centuries in the corpus were observed. Thus a total number of 400 occurrences of participial constructions were evaluated.

Table 4.2 summarises the results of this evaluation. The first column in this table indicates precision results for the 13th century and the second column indicates the same measure for the 20th century. As can be seen from this Table, the results range from 74% to 100% in precision, which guarantees valuable generalizations made on the basis of frequency data. The lowest numbers are obtained for periphrases with *ser* (82%) and *tener* in the 20th century (74%). A certain margin of error is however expected from the fact that the same queries are used throughout the centuries as well as to retrieve all participial constructions.

4.3.2 Design & Procedure

There are some characteristics of corpus data which should be considered in order to draw valid conclusions on the basis of a frequency data. First, the distribution of frequencies across documents in a corpus are often not equally distributed. For example, we could find a high number of occurrences of *estar* + PTCP in one single document whereas we can find none in a document from the same century. It is a well-known fact that a corpus formed from different documents is not a random sample. Thus, it is necessary to take this into account in order to use adequate tests to check for significant changes in the frequencies of corpus data (Church, 2000; Kilgarriff, 2005; Gries, 2006; Evert, 2006). This challenge is even greater for quantitative studies based on diachronic corpora, as these contain texts from different historical periods with very diverse characteristics. For example, as described in chapter 2, the diachronic corpus of Spanish used in this dissertation contains documents of different genres and styles, which may also show

N	Word order	Between Num	Between Class	Right boundary	Left boundary
1	Aux + PTCP	0	-	-	-
2	Aux + PTCP	1	CS;DP;DI;DA;P;I;Z;N;R;A	No mente;Aux	-
3	Aux + PTCP	1	S	No mente;N	-
4	Aux + PTCP	1	CC;F	R*+F,C,PR,S,D	-
5	Aux + PTCP	2	N+A,R,P,I,N; P+R; A+R,N; D+Z,N; CS+A	No mente	-
6	Aux + PTCP	2	N+SP; P+D,N; D,Z+A	No mente;N	-
7	Aux + PTCP	2	N+CC,Z	No mente,Aux,RN	-
8	Aux + PTCP	2	P+CC,F	No mente;V	-
9	Aux + PTCP	2	P+A; A+CC; R+D,N,C,F,R; CS+R;CC+A,P; CC+R,N,CC,F; S+A,R,C,I,S,Z	No mente words;Aux	-
10	Aux + PTCP	2	P+S; R+P; CS+I	-	-
11	Aux + PTCP	2	D,Z+R,S,Z,CS; R+A,S	No mente;Aux;N	-
12	Aux + PTCP	3-7	Any token but V	No mente;PP;N;Aux	-
13	Aux + PTCP	8-10	Any token but V,CS,PR and .	No mente;PP;N;Aux	-
14	Aux + PTCP	11-19	Any token but V.,CS,PR,F	No mente;PP;N;Aux	-
15	PTCP + Aux	0-6	R;PP;PI;PR;CS; Z;I;S; NP	S;CC;CS;NP; A+N;R;F;P;Z;V.[ISM]	C,F,PR,R+NP+Aux

Table 4.1: Queries used to retrieve participial constructions from the corpus.

	13th c.	20th c.	Mean
estar + PTCP	0.86	0.96	0.91
haber + PTCP	0.88	1	0.94
ser + PTCP	0.96	0.82	0.89
tener + PTCP	0.94	0.74	0.84

Table 4.2: Precision of the queries.

significant differences in the use of different grammatical categories.

Partial solutions to this problem include the use of a number of dispersion measures to account for the non-random distribution of data (Gries, 2008), the use of statistical tests for significance which do not assume a binomial distribution, and the study of frequency changes per genre or register.

Here I will adopt a solution which combines the already proposed solutions. In order to explore the dispersion of data and the existence of a general trend of change visually, I will plot relative frequency in each document against time (of publication). Then, for the quantitative analysis of the frequency data I use state-of-the-art techniques from applied statistics. In particular, changes in the usage frequency of the different participial constructions are tested for significance with generalised linear models (GLMs) (Frank E. Harrell, 2001; Baayen, 2008).

The major difference between GLMs and standard linear models (and ANOVA) is that the former is based on the appropriate binomial sampling distribution, specifically it accounts for binomial sampling variation of observed frequencies and different sample sizes , whereas standard linear models approximate it by a Gaussian distribution and assume equal variance, which leads to problems if there are low-frequency items in the sample.

The advantage of GLMs over frequency comparison with traditional chi-squared or Fisher's test is that they do not treat the corpus as a single random sample but as a collection of texts (which is one of the major sources of non-randomness, as explained by Evert (2006)), and that they can account for systematic trends across time (rather than comparing pooled data from different periods, assuming that the relevant frequency is constant across each period).

4.4 Changes in the frequency of participial constructions

As argued before, if there is a change in participial constructions towards greater regularization, this change is expected to be reflected in the frequencies of use of these periphrases. If change is towards greater functional meaning and use, then

we expect to find a significant increase in the frequency; conversely, if meanings associated with a participial construction are eventually lost then we expect to find a significant decrease in its frequency over time.

In order to track changes in the frequency of participial constructions over time, the relative frequency of every participial construction for each document in the corpus was calculated. The relative frequency results from dividing the total amount of occurrences of each one of the constructions by the sum of all the instances of constructions involving *haber*, *ser*, *estar* and *tener* plus past participle. For example, in document *Poema de Mio Cid* (CID) there are 213 instances of *ser* + PTCP. As the total number of participial constructions, formed with *ser*, *haber*, *estar* and *tener*, in this document is 344, then the relative frequency of *ser* + PTCP is obtained by dividing 213 by 344, which is 0.62. The main purpose of calculating per-document relative frequencies is to enable valid statistical inference.

The results of the frequency study from the 12th to the 20th century for each participial construction are illustrated in Figure 4.1. Each point in the graphs corresponds to a single text from the corpus, showing time of composition on the x-axis and the relative frequency of the corresponding construction on the y-axis. As mentioned in chapter 2, texts in *Lexesp* corpus do not have a specific date of composition. Hence, dates (x-axis) in the Late Contemporary period were randomized in order to plot relative frequencies from the part of the corpus coming from the *Lexesp* collection. The solid line in each of these graphs is a locally smoothed average which highlights systematic trends hidden behind the random variation. This line was obtained using the lowess smoother function (available in R), which uses locally-weighted polynomial regression (Cleveland, 1981).

As can be seen from the top left graph in Figure 4.1 the frequency of *haber* + PTCP increases dramatically from the 12th to the 20th century. This is expected from the change undergone by *haber* + PTCP from a stative possessive to a perfect interpretation (greater functional meaning). Interestingly, this periphrasis is changing in 20th century as well. I would suggest that this may be an indication for the change towards the perfective, which has been (qualitatively) observed in some varieties of Spanish, such as Alicante Spanish (Schwenter, 1994). Conversely, as can be seen from the top right graph, the frequency of *ser* + PTCP decreases dramatically over time. This is also expected under my analysis of change of *ser* + PTCP, as this periphrasis loses both its perfect and adjectival passive interpretations. Changes in the frequencies of *haber* + PTCP and *ser* + PTCP are highly significant (Generalized Linear Model with binomial family and logit link, $p < .001$).

Changes of *tener* + PTCP and *estar* + PTCP can not so easily seen from these graphs, as the relative frequencies of these periphrases are considerably lower than those of *haber* + PTCP and *ser* + PTCP. Specifically, the frequency of *tener* + PTCP slightly increases between the 15th and 17th centuries. This can be seen

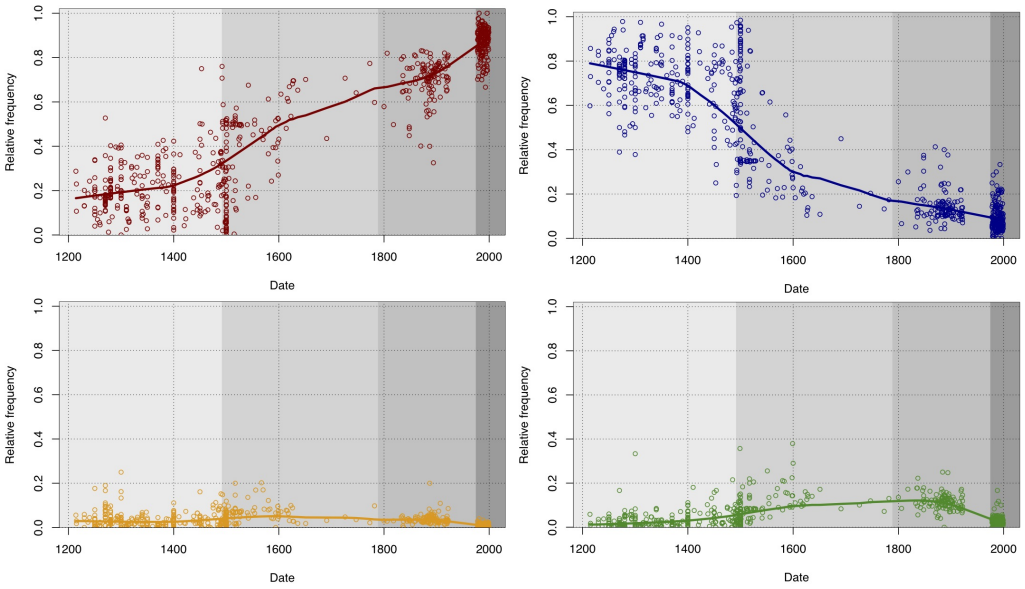

Figure 4.1: Relative frequencies of *haber* + participle (top left), *ser* + participle (top right), *tener* + participle (bottom left) and *estar* + participle (bottom right) from the 12th to the 20th century.

from the bottom left graph in Figure 4.1. Similarly, as can be seen from the bottom right graph, the frequency of *estar* + PTCP increases during the mid centuries and then it decreases during the 20th century. Despite appearances, however, changes in the frequencies of *tener* + PTCP and *estar* + PTCP are also highly significant (Generalized Linear Model with binomial family and logit link, $p < .001$) for both participial constructions. The lower frequencies of *tener* + PTCP and *estar* + PTCP throughout the centuries suggest that the competition for the perfect and verbal passive interpretations would be rather minimal, though statistically significant.

The boxplots in Figure 4.2 compare pooled data for the four periods considered in the corpus, as described in chapter 2. Boxplots facilitate the statistical analysis of the data, as the size of the boxes gives a general impression of the dispersion of the frequency data. The lines in the boxplots show the median for each period.

As can be clearly observed from these plots, two clear stages can be distinguished for changes in the usage frequency of *haber* + PTCP and *ser* + PTCP. The first period shows a slight increase/decrease during the Middle and Modern periods and then the change accelerates from the Modern to the Late Contemporary period. This may suggest that the input probabilities to which speakers were

exposed was also quite variable during that time.

On the other hand, the frequency of periphrases with *tener* and *estar* increases slightly between the Middle and the Modern period and in the Middle and Contemporary period, and then it decreases first in the Contemporary period for *tener* + PTCP and in the Late Contemporary period for *estar* + PTCP.

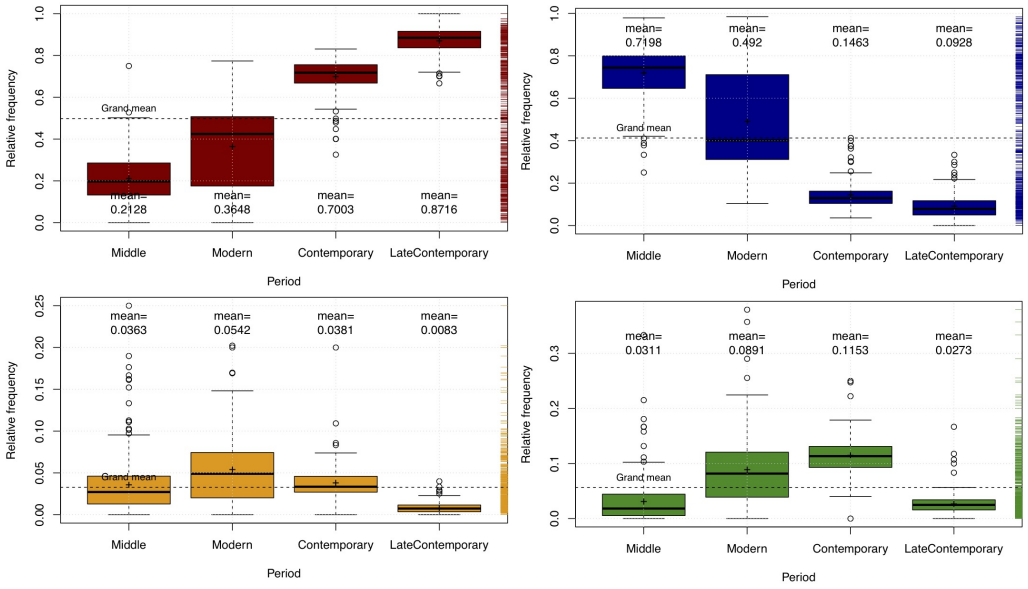

Figure 4.2: Relative frequencies of *haber* + participle (top left), *ser* + participle (top right), *tener* + participle (bottom left) and *estar* + participle (bottom right) in the Middle, Modern, Contemporary and Late Contemporary periods.

Interestingly, from these graphs it can be seen that the distribution of the relative frequencies tends to be concentrated around the means in the latest centuries, showing a more compact dispersion than in the earliest periods. Specially, during the Modern period, the distribution of the frequencies around the mean is quite wide for each construction, whereas the distribution of relative frequencies is much more compact in the Contemporary and Late Contemporary periods. On the other hand, it is important to note as well that the average mean of the relative frequencies of periphrases for each period are much more distant from each other in the Late Contemporary period than in the Modern period. Specifically, in the Modern period the mean frequencies are 0.49, 0.35, 0.05 and 0.09 for the periphrases with *ser*, *haber*, *tener* and *estar*, respectively, whereas the mean frequencies in the Late Contemporary period are 0.87, 0.093, 0.008 and 0.027 for the same periphrases. Crucially, in language acquisition studies similar results have been reported for regularization processes, where chains of learners exposed to

different probabilistic inputs tend to converge after a few generations towards a range of frequencies in the choice of grammatical markers (Smith and Wonnacott, 2010).

All in all, I interpret changes in the frequency of participial constructions, as just presented, as evidence for the regularization change that they were undergoing, by which they became the expressions for only one interpretation in an univocal way.

4.5 Concluding remarks

In this chapter I have provided quantitative evidence for the regularization change undergone by participial constructions in Spanish, as proposed in chapter 3.

Furthermore, based on these observations it is also possible to propose a chronology for change, in which there is a steady change until the 14th century approximately, and then the change accelerates up to the 18th and 19th centuries for all participial constructions. Also we have observed that first *tener* + PTCP and then *estar* + PTCP reach a certain stabilisation after the Modern period, and, conversely, that the development of periphrases with *haber* continue until the Late Contemporary period. In the case of *ser* + PTCP we have seen a steady decrease in it frequency over time, which is probably an indication that this periphrasis is loosing some of its functions.

Ultimately, this study provides relevant empirical evidence which comes from real corpus data to explore regularization processes of change as they happen over long periods of time.

5

THE LEXICAL EXTENSION OF CHANGE

5.1 Introduction

In the previous chapter I argued that participial constructions display different yet connected regularization changes, based on the observations of changes in their frequency distributions in the corpus. Specifically, we saw how the frequencies of these periphrases tend to change towards clearly distinguishable frequency distributions, in which one construction has become the vast majority (*haber* + PTCP) whilst the others are less frequent, and where the distribution of the relative frequencies tends to be concentrated around the means, showing a more compact dispersion than in the earliest periods, when variation was more complex.

I interpreted these findings as evidence for a competition between these constructions to express the same interpretation and the subsequent regularization change, by which each construction eventually comes to express one interpretation. The system changes from being very complex, as each periphrasis can express more than two interpretations and each interpretation can be expressed by more than one periphrasis, to a one-to-one form-meaning system. In the course of this change the periphrasis with *ser* loses the perfect and adjectival interpretation over time, and that with *tener* the possibility to instantiate perfects. In contrast, *haber* and *estar* become the only grammatical markers for perfects and adjectival passives, while they lose their stative possessive and verbal passive interpretations respectively. These stative possessive and verbal passive interpretations eventually come to be expressed only by *tener* + PTCP and *ser* + PTCP. Furthermore, I suggested that the increase in the frequency of *haber* is expected from grammaticalization changes, as these are changes towards a more functional use of a given expression in languages.

This chapter addresses the following question: Which is the driving force behind the changes undergone by these participial constructions? More specifically, how do *haber* and *estar* extend as the only auxiliaries for perfects and adjectival passives? And, conversely, how do *ser*, *estar* and *tener* cease to express perfects or verbal passives? Do these periphrases develop these uses in all contexts at the same time or does this change takes place first in some contexts and later in others?

The goal of this chapter is to begin to answer these questions, by investigating, on the one hand, the spread of *haber* + PTCP as the perfect, on the other, the spread of adjectival passives with *estar* + PTCP. The loss of the perfect and adjectival passive interpretations of *tener* + PTCP and *ser* + PTCP will be explored concomitantly, based on the assumption (supported by the data) that this change is the result of the competition between all four periphrases.

The main idea to be explored in this chapter, on the basis of a quantitative study of changes in auxiliary-verb combinations, is that change in these periphrases spreads gradually through the lexicon, from some semantic classes of verbs to others. In this chapter I will show how the frequency with which each auxiliary combines with a particular class of verbs influences the subsequent development, in a way which can be explained by general cognitive mechanisms such as priming. In this respect, I will try to go one step further than previous accounts, e.g Mendeloff (1964), Pountain (1985), or Batllori *et al.* (2009) and Batllori (2011), which restrict the study of the development of these constructions to the morphosyntactic manifestations of the change such as participial agreement and word order changes.

The results of the empirical investigations, as I will argue in what follows, suggest that the development of participial constructions in Spanish is sensitive to the lexical semantics of the base predicates, in ways which differ in the case of the spread of the perfect with *haber* and of adjectival passives with *estar*.

The structure of this chapter is as follows. In Section §5.2 I introduce the notion of lexical extension. Sections §5.3 and §5.4 are the central sections in this chapter, where I explore the extension of perfects with *haber* and adjectival passives with *estar* and explore the idea, based on quantitative evidence, that change spreads through the lexicon. I conclude this chapter in Section §5.5 with a summary and a general discussion about change in participial constructions.

5.2 Lexical extension and priming

Lexical extension (also called **lexical diffusion**) is the term used to refer to a type of extension that happens gradually through the lexicon, one or several words at a time. In the literature this notion has often been used to explain how sound

change can happen in language, in that it is phonetically abrupt but lexically gradual or spreading word by word through the lexicon (Wang, 1969; Wang and Chin-Chuang, 1977; Labov, 1981).

Lexical extension has also been used to explain how some syntactic changes extend in time. For example, Joseph (1983) documents the lexical extension in the loss of the infinitive in Greek and other Balkan languages. In Classical Greek a large number of verbs governed infinitives in a variety of contexts, and these were gradually replaced with finite verb forms with a particle or complementizer.

In this chapter I investigate how perfects expressed by *haber* + PTCP and adjectival passives by *estar* + PTCP gradually spread through the lexicon at different rates with some types of verbs than with others, and how this extension takes place at the expense of the other participial constructions which were competing to express the same interpretations.

Another idea that will be explored here, specifically in the case study of the lexical extension of adjectival passives, is that the original context of use of auxiliary verb may eventually determine the subsequent development of the participial construction formed with the given auxiliary. I suggest that the mechanism which underlies such development is **priming**, whose role in language production, as well as in language comprehension and dialogue has largely been attested (Bock and Kroch, 1989; Luka and Barsalou, 2005; Pickering and Garrod, 2004). Priming is preactivation in the sense that the previous use of a certain linguistic element (or **prime**) will affect the subsequent use of the same or a sufficiently similar element (or **target**). In this respect, the development of participial constructions in Spanish is evidence for a view of language change in which both lexical semantics of the linguistic expression(s) in context as well as the frequency or repetition of certain meaning-form pairs play an important role.

5.3 The development of the Spanish perfect through aspectual classes of predicates

5.3.1 Introduction

As it was established in the previous chapter on the basis of quantitative data, it takes many centuries for *haber* + PTCP to become the sole expression of the perfect as it is today. Occurrences of this participial construction can already be found in Ciceronian Latin and there is evidence that it is still changing until the Late Contemporary period.

As mentioned before, in the earliest centuries *haber* + PTCP was used as a stative possessive construction roughly describing a relation of possession between the subject and the internal argument, which could be further modified by a sec-

ondary predicate. This secondary predicate could be a participle, as well as other categories such as adjectives, or prepositional phrases. The meaning of this construction could be roughly paraphrased as *X have Y in Z state*, where X is the possessor subject, Y is the possessed entity, coded as the internal argument, and the secondary predicate Z, which describes a stative property of the internal argument. Old Spanish examples in (1), both unacceptable in Modern Spanish, illustrate this. (1-b) is especially clear in this respect, as we see the same syntactic frame for the paraphrased meaning proposed above.

(1) a. *E quando las uieren tristes e espeluzradas. sepan que* ***an***
and when them saw sad and frightened know that have
alguna enfermedad ***encubierta***.
some illness hidden
"And when they saw them sad and frigthened, you must know that they have some illness hidden" (13th c., MOA)

b. *E este anjuersario fizo sienpre el rrey don alfonso cada año*
and this anniversary made always the king Mr Alfonso each year
en quanto ***ouo*** *los Regnos* ***en su poder***
in how had the kingdoms in his power
"And the King Alfonso celebrated this anniversary every year after he had the kingdoms in his power" (13th c., CAX)

Based on the qualitative analysis of corpus data, in chapter 3 I suggested that the perfect interpretation to *haber* could be derived from this stative possessive interpretation by a (gradual) process in which an unspecified two-place predicate *haber* becomes an auxiliary, and together with the participle they form a compound predicate, where the external argument of the eventuality described by the participle is coded as the subject of the auxiliary. The findings in the present study thus seem to confirm the traditional view that current *haber* + PTCP perfect constructions emerge from what I call here the stative possessive construction.

The question may now arise as to how the perfect developed in time, so that it came to be expressed only by *haber* + PTCP. From the language change perspective, one possible way to explore the extension of perfects with *haber* would be to inspect those contexts where the possessive interpretation of *haber* is not possible. In these contexts the subject of *haber* should be clearly interpreted as the external argument (subject) of the participle as well.

As possessive uses of *haber* are transitive, in that the internal argument (possessee) is coded as direct object, such contexts would be those in which *haber* is used intransitively. In these cases, *haber* should appear in combination with truly intransitive predicates (that is unergatives), as only this type of predicates does not have an internal argument which could in turn be interpreted as the object of

haber. There are, however, some problems with this approach. First, it is difficult to conduct a quantitative study of this, given the variable syntactic behavior of many verbs, which may behave as transitives in some contexts and as intransitives in others.

Furthermore, it is difficult to track changes in the frequencies of auxiliaries in combination with intransitives if we do not have syntactically annotated corpora. For this reason, here I have pursued an alternative strategy, which consists of exploring the development of the perfect through aspectual (or Aktionsart) classes of predicates, specifically accomplishments, achievements, states and activities. Instead of studying changes in the syntactic type of the verbs combining with *haber* and its gradual extension to all syntactic types of predicates, here I track changes in auxiliary-participle combinations with different semantic classes of verbs. From the language change point of view, tracing the development of auxiliaries with these verbs will allow us to see which are the semantic features relevant to this change.

An additional advantage of this strategy is that we will be able to keep track of the spread of the different perfect readings. As presented in chapter 3 on the basis of previous accounts of the perfect category (Dowty, 1979; Mittwoch, 1988; Vlach, 1993; Portner, 2003; Iatridou *et al.*, 2003, among many others), particular readings are available to the perfect depending on the type of predicate the auxiliary combines with. Specifically, individual-level (IL) states can yield both U- and E-perfect readings, as in English sentences in (2-a) and (2-b) (also shown in chapter 3, but repeated here for convenience); whereas activities can only yield E-perfects, as in sentence (2-c).

(2) a. Mihajlo **has been** in Barcelona since Tuesday.
b. **Have** you ever **been** to Zaragoza?
c. John **has smoked**.
d. Berit **has arrived** in Paris.

This fact contrasts with R-perfects, as in (2-d), which can only be formed of those verbs which lexically encode a consequent state, that is telic eventualities. In Aktionsart terms, this type of predicates are accomplishments and achievements. While base verbs in E-perfects can be both accomplishments and achievements, U-perfects cannot be built from any of these types of verbs. This is summarised in Table 5.1.

On the other hand, in Spanish (quite generally) there seems to be some correspondence between the aspectual semantic class and the syntactic type of the verb; specifically, there are many intransitives among activities, whereas transitives and unaccussatives are often accomplishments, achievements or IL-states. Hence, ultimately we are indirectly keeping track of the syntactic type of predicates as well,

	U-Perfect	E-Perfect	R-Perfect
Auxiliary + Activity	-	+	-
Auxiliary + IL-State	+	+	-
Auxiliary + Accomplishment	-	+	+
Auxiliary + Achievement	-	+	+

Table 5.1: Perfect readings conveyed by aspectual classes of predicates.

thus providing an indirect way to keep track of syntax as well.

As the same restrictions (over the type of the base predicate) apply to resultative perfects and stative passives and possessives, the development of the perfect can mainly be traced by exploring changes in auxiliaries with activities and IL-states. Thus, from the language change point of view, when the auxiliary is combined with accomplishments and achievements the construction could be interpreted as well as a stative possessive or an adjectival passive. In quantitative terms, as *haber* + PTCP is the succeeding perfect form, we would thus expect to find an increase in the frequency of *haber* plus IL-states and activities. Conversely, the loss of the perfect readings of *ser* + PTCP, *estar* + PTCP and *tener* + PTCP should lead to a decrease in its frequency with states and activities, both of which may yield universal and existential perfects.

5.3.2 Data & Verb classification

In order to conduct this study, all combinations of *haber*, *tener*, *estar* and *ser* with aspectual classes of predicates were obtained from the corpus using *Corpus Workbench*. The same query syntax used to obtain all counts of participial constructions, as described in chapter 4, was used in this study as well.

Lists of aspectual classes of verbs in Spanish were carefully compiled by Rafael Marín and myself. There are 50 verb types forming each class, categorized on the basis of a number of linguistic diagnoses. The diagnostics which distinguish aspectual classes of verbs have been extensively discussed in the literature; see Vendler (1957); Dowty (1979) and specifically for Spanish de Miguel (1999); Marín (2000). These tests are sensitive to a number of features of the meaning of these verbs such as agentivity, telicity or dynamicity. The complete list of verbs is in Appendix C, but I will reproduce here some examples for illustration:

- IL-states, e.g. *abundar* 'be abound', *admirar* 'admire', *adorar* 'adore'...
- Activities, e.g. *acariciar* 'caress', *andar* 'walk', *bailar* 'dance'...
- Achievements, e.g. *abrir* 'open', *acertar* 'get right', *adquirir* 'acquire',...

- Accomplishments, e.g. *arrasar* 'devastate', *arreglar* 'fix', *ascender* 'ascend',...

In order to guarantee that these verbs appear in the corpus, we included in the classification only those verbs which appear in the 20th-century part of the corpus more than 15 times. On the other hand, in this study verbs with an extremely high frequency in participial constructions, such as *decir* and *hacer*, have been excluded from the classification, in order to exclude potential outliers from the statistical analysis, which may bias the results.

The total number of auxiliaries with these aspectual classes of verbs accounts approximately for between 15% and 20% of the total number of participial constructions throughout the centuries, which is sufficient to get an impression of the general tendency.

Before we move on to present the results, a cautious note should be mentioned here about the class of IL-states. As it has been extensively discussed in the literature, there are different types of stative predicates. For example, Dowty (1979) mentions instantaneous and non-instantenous states, Bach (1986) talks about dynamic and non-dynamic states. In Spanish, Marín (2000) proposes to distinguish between bounded and unbounded states. In this study IL states correspond to unbounded states as in Marín (2000). These predicates are characterised, for example, by their incompatibility to be the complement of verbs like *parar* or *dejar* 'stop'. As it will be shown later, other types of stative predicates show different patterns than IL-states in participial constructions, and therefore if one wants to keep track of this over time, it seems methodologically adequate to treat these types separately.

5.3.3 Results

The graphs in Figure 5.1 illustrate the results of the corpus study. Tables 5.2, 5.3, 5.4, 5.5 indicate the relative frequencies and total numbers of each auxiliary in combination with each aspectual class of verb which are the basis for these graphs.[1] Each line in the graphs in 5.1 illustrates the development of auxiliaries in combination with aspectual predicates (as indicated in the legend), showing time on the x-axis and relative frequency on the y-axis. Relative frequencies are here calculated over the total number of participial constructions combining with each aspectual class of verbs. For example, given that the total number of *ser* plus IL-states in the 13th century is 4637, and that the total number of *ser*, *estar*, *haber* and *tener* with IL-states in this century is 5217, then the relative frequency of *ser* with IL-states in this century is 0.89 (4637 divided by 5217).

[1]The total number of auxiliaries with each aspectual class of verb per century is given in Appendix E.

As can be seen from the top left graph, the frequency of *haber* increases gradually with all aspectual classes of predicates since the earliest centuries, especially after the 15th century. Interestingly, the proportions of *haber* with all classes tend to converge in the 20th century, when percentages range from 85% to 90% approximately, clearly suggesting the general spread of *haber* with these predicates over the other auxiliaries. The increase in the frequency of use of *haber* with all these verbs is expected from the grammaticalization of *haber* + PTCP as the only expression for the perfect interpretation. It is noticeable that the frequency of *haber* with activities is higher than with the other auxiliaries already from the earliest centuries.

Conversely, as can be seen from the top right graph in Figure 5.1, the frequency of *ser* gradually decreases with all aspectual classes of verbs since the earliest centuries. This is also expected from the fact that *ser* + PTCP eventually ceases to express the perfect interpretation. A similar general pattern for change can be observed with *haber* + PTCP, but in an inverse way: change begins slowly with small shifts in the proportions in the earliest centuries, mostly with activities and achievements but also with accomplishments and states. The decrease in the frequencies speeds up from the 15th century onwards, reaching a stage in the 20th century when they are used in similar proportions (from 11 to 4 per cent).

Moving to the other participial constructions, we also observe interesting facts. As can be seen from the bottom graphs, the proportion of these auxiliaries with all aspectual classes of verbs slightly increases between the 15th and 19th century and then it decreases, yielding the same (quite low) proportions by the 20th century. In the earliest centuries the proportion of *tener* and *estar* with activities is almost totally split between *ser* and *haber*, leaving only the remaining 3% to *estar* and *tener*. The frequency changes very slighly in the case of activities and IL-states, which yield very low relative frequencies with these auxiliaries in the 20th century. This is also expected from the assumption that periphrases with *tener* and *estar* were competing to express the perfect interpretation in Old Spanish, as combinations of auxiliaries with activities and IL-states can only give rise to this interpretation. I take this as an indication of the loss of the perfect interpretation with participial constructions formed with these verbs. These data would also indicate, as suggested in the previous chapter, that the competition of these constructions for the perfect interpretations was minimal, in comparison with *ser*.

5.3.4 Discussion

To recap, by exploring changes in frequency of auxiliaries with aspectual classes of verbs we have seen more specifically how *haber* + PTCP spreads as a perfect, taking over the perfect uses of the other periphrases. However, the fact that the frequency of *haber* with activities is quite high already in the 13th century suggests

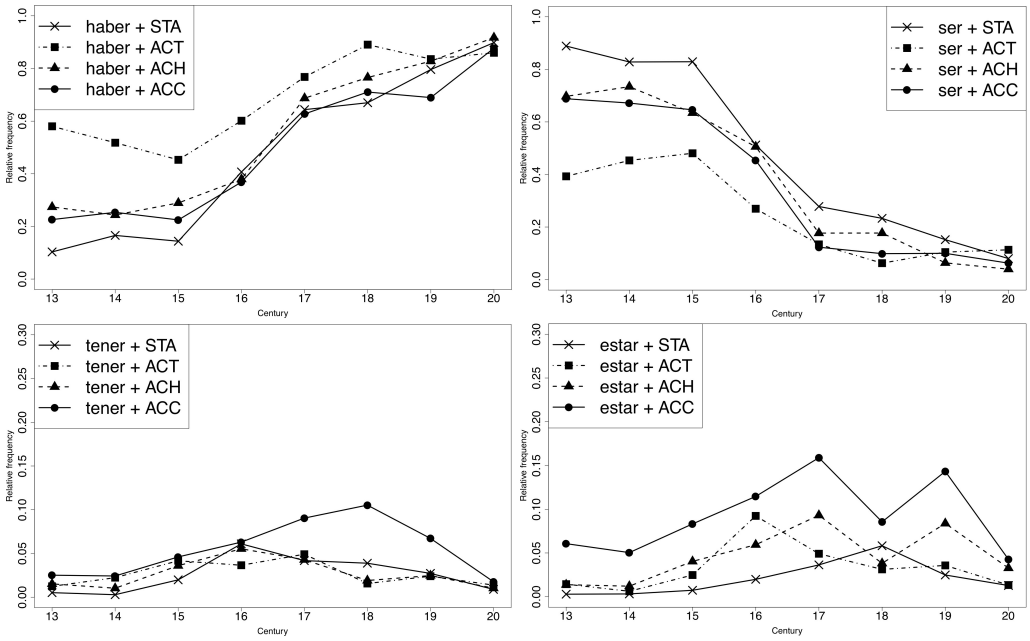

Figure 5.1: Frequency of *haber* (top left), *ser* (top right), *tener* (bottom left) and *estar* (bottom right) plus aspectual classes of predicates from the 13th to the 20th century. In the legends 'STA' stands for IL-states, 'ACT' for activities, 'ACC' for accomplishments and 'ACH' for achievements.

	%haber	*%ser*	*%estar*	*%tener*	Total
13th c.	10.3	88.9	0.3	0.5	5217
14th c.	16.6	82.8	0.3	0.3	2823
15th c.	14.4	82.9	0.7	2.0	5045
16th c.	40.7	51.2	2.0	6.1	2214
17th c.	64.4	27.8	3.6	4.2	550
18th c.	67.0	23.3	5.8	3.9	206
19th c.	79.6	15.2	2.5	2.7	4995
20th c.	89.8	8.0	1.3	0.9	4378

Table 5.2: Relative frequency of IL-states in participial constructions with *haber*, *ser*, *estar* and *tener* from the 13th to the 20th century.

that the participial construction formed with this auxiliary was already frequently used to express perfects since the earliest centuries.

As presented in chapter 1, in earlier literature it was proposed that the loss of agreement is one of the facts leading to the grammaticalization of the perfect with

	%haber	*%ser*	*%estar*	*%tener*	Total
13th c.	58.0	39.3	1.4	1.2	1046
14th c.	51.8	45.4	0.6	2.2	807
15th c.	45.3	48.1	2.5	4.2	1009
16th c.	60.2	27.0	9.2	3.7	931
17th c.	76.8	13.4	4.9	4.9	224
18th c.	89.1	6.3	3.1	1.6	64
19th c.	83.6	10.5	3.6	2.4	2209
20th c.	86.0	11.3	1.3	1.4	2251

Table 5.3: Relative frequency of activities in participial constructions with *haber*, *ser*, *estar* and *tener* from the 13th to the 20th century.

	%haber	*%ser*	*%estar*	*%tener*	Total
13th c.	22.6	68.8	6.1	2.5	3182
14th c.	25.4	67.2	5.0	2.4	2063
15th c.	22.5	64.6	8.3	4.6	2608
16th c.	36.8	45.4	11.5	6.3	1509
17th c.	62.7	12.3	15.9	9.0	365
18th c.	71.1	9.9	8.6	10.5	152
19th c.	69.0	10.0	14.3	6.7	2960
20th c.	87.7	6.3	4.3	1.8	3230

Table 5.4: Relative frequency of accomplishments in participial constructions with *haber*, *ser*, *estar* and *tener* from the 13th to the 20th century.

	%haber	*%ser*	*%estar*	*%tener*	Total
13th c.	27.4	69.7	1.4	1.5	10339
14th c.	24.3	73.4	1.2	1.0	7316
15th c.	28.9	63.4	4.0	3.6	7553
16th c.	37.9	50.6	5.9	5.5	4917
17th c.	68.7	17.7	9.3	4.3	1074
18th c.	76.6	17.7	3.8	1.9	367
19th c.	82.8	6.4	8.4	2.5	9645
20th c.	91.7	4.0	3.3	1.0	10595

Table 5.5: Relative frequency of achievements in participial constructions with *haber*, *ser*, *estar* and *tener* from the 13th to the 20th century.

haber + PTCP (Romani, 2006). However, the findings about the lexical extension of the perfect with *haber* through aspectual classes of verbs, as just presented,

suggest that changes in the semantic interpretation of this periphrasis may have led to change, and that morphological changes in the periphrasis are a subsequent development, and hence not the cause but rather the surface manifestation of the underlying change in the interpretation of the construction.

From the language change point of view, by exploring changes in gender and number agreement in the given constructions we could track the spread of perfect with *haber*, as opposed to the other periphrases. Quantitatively, if the lack of agreement with the internal argument is one of the morphosyntactic characteristics of the Modern Spanish perfect, we expect to find a decrease in the frequency of occurrences of *haber* plus agreeing participles over time. In contrast, we do not expect to find this decrease with the other participial constructions, as they do not become the perfect but eventually come to express passives and statives.

In order to check this idea in the corpus, all participles combining with auxiliaries in masculine plural or feminine plural and singular forms were retrieved from the corpus. As a cover term I will refer to this type of participles as **agreeing**. Participles with masculine singular morphology (*-o*) are not considered in this study, as this form is also the neutral ending which is used in the Modern Spanish perfect participle. There is therefore no way to differentiate cases of agreeing from non-agreeing perfect participles with *-o* in the corpus during the (automatic) retrieval process.

Table 5.6 summarises the corpus findings for all participial constructions. Figure in 5.2 illustrates this with a graph. Columns headed by % indicate the percentage points of participles in masculine plural, feminine plural and singular forms with respect to the total number of occurrences of each construction, which is indicated in the columns headed by *Total*.

	Haber + PTCP		*Ser* + PTCP		*Tener* + PTCP		*Estar* + PTCP	
	%	Total	%	Total	%	Total	%	Total
13th c.	24.1	24,236	48.1	65,033	60.5	2,661	62.6	2,484
14th c.	22.2	16,453	49.0	45,605	62.1	1,202	59.8	1,292
15th c.	12.3	18,034	51.3	65,019	58.2	2,569	56.1	3,509
16th c.	5.6	15,579	55.4	32,085	54.0	2,551	61.4	3,708
17th c.	1.3	5,636	47.8	1,838	50.6	599	44.1	1,091
18th c.	1.2	2,239	59.2	537	62.1	124	60.4	331
19th c.	1.3	52,261	53.3	10,107	55.4	2,750	52.5	8,314
20th c.	0.5	58,549	51.7	6,923	56.4	1,082	51.1	3,574

Table 5.6: Percentage of agreeing participles (in masculine plural, feminine singular and plural form) with *haber*, *ser*, *tener* and *estar* from the 13th to the 20th century.

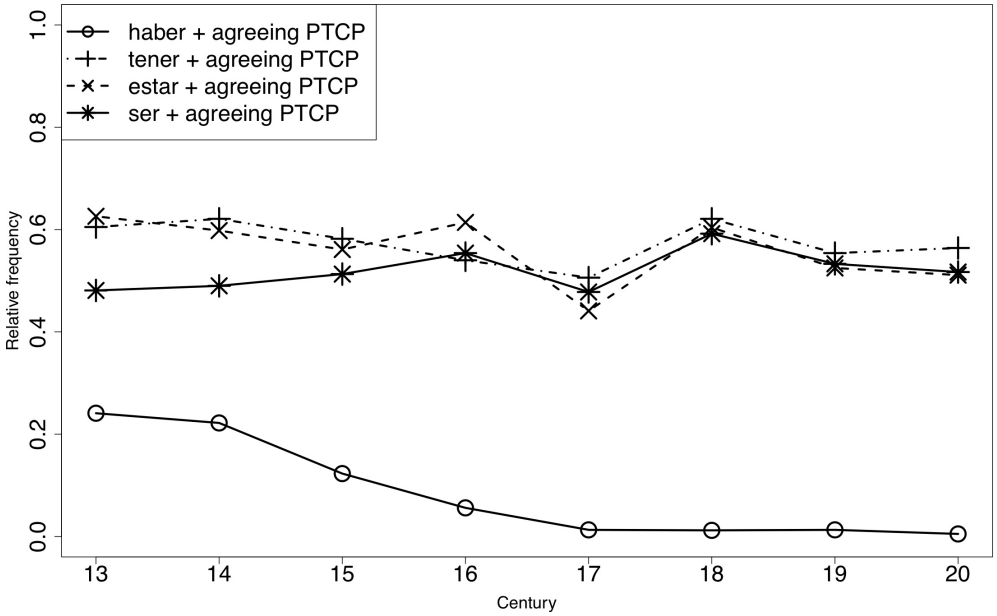

Figure 5.2: Frequency of *haber*, *ser*, *tener* and *estar* plus participles in masculine plural or in feminine singular and plural form from the 13th to the 20th century.

As can be seen from the table and the graph, the proportion of these participles with *haber* decreases gradually from the 14th to the 17th century. The rate of decrease is approximately 50% between the 14th and 17th century. By the 17th century the proportion of agreeing participles becomes relatively stable and in the 20th century agreeing participles are almost non-existent. In contrast, proportions of these participles with *ser* + PTCP, *tener* + PTCP and *estar* + PTCP do not change significantly over time. They range from approximately 50 to 60% throughout the centuries, as expected from the fact that they are the expression for adjectival and verbal passives as well as stative possessives, which always show agreement.

The fact that the proportion of agreeing participles with *haber* in the earliest centuries is already quite low, in comparison with the other auxiliaries, suggests that this construction was already more frequently used as a perfect than the other periphrases since the earliest centuries. In contrast with the other participial constructions, the proportion of agreeing participles with *haber* yields 25 per cent in the 13th century, which is more than 50% lower than the percentage of agreeing participles combining with the other auxiliaries.

5.4 Origins and change of adjectival passives

5.4.1 Introduction

So far we have explored the spread of the perfect with *haber*, and the loss of these uses with *ser*, *tener* and *estar*. However, this is only one side of the coin. During the same time span there was a change in passives as well, which only affected participial constructions formed with the copulas *ser* and *estar*. In the earliest centuries *ser* + PTCP instantiated both adjectival and verbal passives, similarly to Modern English. However, some time later by the 13th century *estar* + PTCP began to be used productively with both passive interpretations as well. At that time then adjectival and verbal passives could be expressed by both constructions. This is illustrated in (3) and (4):

(3) (Verbal passives in Old Spanish)

a. *E dixo otro: saber las cosas ayuda a obrar: & si no pudieres*
and said other know the things helps to operate and if not could
ser amado *sey amador.*
be loved be lover
"And the other said: knowing things helps to do; if you cannot be loved be a lover" (13th c., BDS)

b. *Contra esto dize el otro que mas val* ***estar amado*** *que*
agains this says the other that more worth be.LOC loved than
temjdo
feared
"The other says against this that it is worth more being loved than feared" (13th c., LAT)

(4) (Adjectival passives in Old Spanish)

a. *començando de las yslas que* ***son pobladas*** *en la Mar*
beginning of the islands that are inhabited in the sea
occidental; a qui dizen la Mar occeana.
occidental to who say the sea ocean
"... beginning with the islands that are inhabited, located in the Occidental sea, which they call the Ocean sea" (13th c., ALB)

b. *Et la villa* ***esta poblada*** *en vn otero alto que esta junto*
and the village is.LOC inhabited in a hillock high that is.LOC by
con el mar Et tiene tres çercas & entre cada çerca mora
with the sea and has three sieges and between each siege resides
gente
people

"And the village is inhabited in a high hillock that is located by the sea; and it has three sieges, beside which there are people living" (15th c., TAM)

Eventually *ser* + PTCP comes to express only verbal passives, whereas *estar* + PTCP spreads as the only expression of adjectival passives. As it is well-known, in Modern Spanish verbal and stative passives are morphologically distinct. Verbal passives are expressed by *ser* + PTCP (5-b) whereas stative passives are expressed by *estar* + PTCP (5-a) (in German the copulas used are *werden* and *sein*, respectively).

(5) (Adjectival and verbal passives in Modern Spanish)

a. *El piso que ahora habito no* ***está*** ***arreglado*** *por obreros, sino*
the flat that now live.1SG not is.LOC fixed by workers but
por personas que se hacen pasar por obreros.
by people that REFL make pass for workers
"The flat in which I am living now is not fixed by workers, but by other people who pretend to be workers" (20th c., a28)

b. *Los salones* ***fueron vendidos*** *a los anticuarios de Florencia.*
the saloons were sold to the antique shops in Florence
"The saloons were sold to the antique shops in Florence" (20th c., GHKT)

In this section I focus on studying how *estar* + PTCP emerged in Spanish and how it was generalized in time as the only means to express adjectival passives, eventually winning over the stative uses of the periphrasis with *ser*.

Before we move on, I would like to point out two crucial differences between the development of the perfect, as presented in the previous section, and change in adjectival passives. The first difference concerns the developmental stage of the given participial construction we are studying; and the second is concerned with the typological or geographical extent of periphrases instantiating adjectival passives.

Firstly, in contrast to *ser* and *haber* periphrases, which trace back their origins to Latin (Cennamo, 2008), *estar* + PTCP began to be productively used only later in the 13th century and specifically in the geographical area of the Iberian Peninsula. There are some occurrences of the cognate Latin verb *sto* 'remain, stand' with past participles, but they are still very rare, especially in the Classical and Archaic Latin period. At that time, this predicate was often used either in combination with prepositional phrases to express the location of some entity in space, as in (6-a), or to describe the continuation of some state of affairs, as in (6-b).

(6) a. ***Sto*** *ad ianuam.*
am.1SG in front of door
"I am in front of the door"
b. *Pugna* ***stetit****.*
battle continued
"The battle continued"

The second difference concerns the geographical extent of the *estar* periphrasis among Romance languages. A distinct morphological expression for the adjectival passive exists only in Spanish; in other Romance languages like Italian or French - as mentioned in chapter 1 - adjectival passives are expressed using the same copula that is used in verbal passives.

From a more general perspective, there is yet another difference concerning the way perfects and passives are expressed in Old Spanish. In contrast to perfects, which in Old Spanish could be realized by all four auxiliaries *haber*, *ser*, *estar* and *tener*, adjectival and verbal passive interpretations could only be expressed by *ser* and *estar*. This narrower range of possibilities in the morphosyntactic instantiations for both types of passives in Old Spanish is expected from the fact that in passives the subject encodes the internal argument of the underlying verb, which has a patient role. As old possessive *tener* and *haber* were transitives, their internal argument is always coded in a direct object position and therefore a passive interpretation is not logically possible.

The central claims about the extension of adjectival passives with *estar* from Old to Modern Spanish which I want to argue for are as follows. First, there is some change in *estar*, which in its Latin origins, as indicated before, was only used as a durative state or as a postural verb describing the location of an entity. This change enables the copula to combine with participles in order to form adjectival passives.

In contrast with the perfect development, adjectival passives with *estar* + PTCP spread through the lexicon with some semantic classes of verbs before than others. In this case, however, there is a fundamental difference with respect to the spread of the perfect. Possibly because we are looking at the earliest stages in the development, the lexical semantics of locative prepositional phrases combining with *estar* since Late Latin determine the emergence and subsequent development of the adjectival passive expressed by *estar* + PTCP. The cognitive mechanism that can explain this development is analogy or structure-mapping, in a way which will be discussed in section 5.4.4.

Once the periphrasis with *estar* emerges, there is a competition between *ser* + PTCP and *estar* + PTCP to express verbal and adjectival passives, which eventually leads to the extension of the innovative periphrasis with *estar* as the only means to express adjectival passives, and conversely, to a replacement of *ser* in

these uses, which specializes as a verbal passive.

Hence, there are three basic steps in the development of Spanish passives: First, change in *estar*; second, emergence of the innovative construction *estar* + PTCP, and finally extension of this innovation, as opposed to the older expression for adjectival passives *ser* + PTCP.

5.4.2 Data & Verb classification

In order to explore in the corpus the idea that adjectival passives with *estar* spread through the lexicon, some lists of verbs were carefully compiled. In this case study the participial constructions explored here are all combinations of *ser* and *estar* + PTCP, obtained from the corpus using the same method as explained in the previous section.

For various reasons that will become clear in a moment, classical aspectual classes of predicates are not really useful to explore the spread of adjectival passives with *estar* + PTCP. In contrast with the analysis of the development of the perfect, as presented in the previous section, by only looking at change in copula combinations with aspectual classes we cannot really understand the development of passives. As both passives can be built from accomplishments and achievements, both of which encode a consequent state in their meaning, and conversely neither of these passives appears very frequently with IL-states and activities, we cannot really trace changes in passives by exploring changes in these copula-participle combinations.

On the other hand, as it has been largely noted in theoretical studies, there are certain types of verbs which do not fit into classical aspectual classifications, such as degree achievements or psychological predicates. For example, it has been acknowledged on multiple occasions that some English verbs show variable aspectual behavior with respect to standard telicity tests, such as the compatibility with time-span adverbials (Krifka, 1989; Tenny, 1994; Levin and Rappaport, 1995; Hay *et al.*, 1999; Piñón, 2005, among many others). Verbs of creation and motion like *drink* (7) and *run* (8), and degree achievements like *widen* (9) give rise to a telic interpretation in some contexts, as in (7-a),(8-a) and (9-a), whereas in other contexts they behave like atelic predicates, as in (7-b), (8-b), and (9-b).

(7) a. John drank a glass of wine in an hour.
b. John drank wine for an hour.

(8) a. John ran the marathon in an hour.
b. John ran for an hour.

(9) a. The canyon widened 30 kilometers in one million years.
b. The canyon widened for one million years.

These verb types pose a challenge for traditional aspectual classes of predicates such as states, activities, accomplishments and achievements, which are mostly diagnosed precisely on the basis of their temporal properties (Vendler, 1957; Dowty, 1979). Crucially, some of these verbs show important contrasts in their behavior with both types of passives in Modern Spanish, and therefore it would be useful to explore how passives have developed in time with these classes of verbs.

For example, there is a number of verbs that, despite being transitive, are odd with *ser* in some contexts. This is the case of some degree achievements like *acortar* 'shorten', *desarrollar* 'develop', or *oscurecer* 'darken', object experiencer psychological verbs such as *apasionar* 'have a passion', *obsesionar* 'get obsessed', and some locative verbs like *rodear* 'surround' with inanimate subjects. See some examples of these verbs from Modern Spanish in (10).

(10) a. ??*El pantalón* ***fue*** ***acortado*** *por su madre.*
the trousers was shortened by his mother
b. ??*La fachada de la casa* ***fue*** ***oscurecida.***
the facade of the house was darkened
c. ??*María* ***fue*** ***obsesionada*** *por Juan.*
María was obsessed by Juan
d. ??*La ciudad* ***fue*** ***rodeada*** *por murallas.*
the city was surrounded by walls

Furthermore, some types of accomplishments and achievements tend to appear preferably with adjectival passives. These verbs include so-called **location** and **locatum** verbs like *encarcelar* 'jail' or *engrasar* 'grease', respectively (Levin, 1993; Buck, 1993). Also incremental change verbs like *comer* 'eat', can appear forming adjectival passives, but are a bit rare in some contexts. See examples of this in (11):

(11) a. *Los asesinos* ***están*** ***encarcelados.***
the killers are.LOC imprisoned
b. ??*Los asesinos* ***fueron*** ***encarcelados*** *por la policía.*
the killers were.LOC imprisoned by the police
c. ??*La manzana* ***está*** ***comida.***
the apple is.LOC eaten

In order to explore the lexical extension of *estar* + PTCP through verb types, I obtained all occurrences of *ser* and *estar* with these types of verbs from the corpus. To do this I worked out a list with these types of verbs. The full lists of verbs can be found in Appendix D. Specifically, the following classes have been

distinguished (the number in parentheses indicates the number of verbs in each class):

- Degree achievements (100), like *congelar* 'freeze' and *alargar* 'lengthen'
- Object experiencer psychological verbs (100 verbs), such as *enfadar* 'get angry' and *asustar* 'frighten'
- Incremental change verbs (25 verbs), e.g. *comer* 'eat', *crear* 'create'...
- Extent predicates (20 verbs), e.g. *extender* 'extend', *cubrir* 'cover', *esparcir* 'scatter'...
- Location and locatum verbs (I will use the term **locative** as a cover label for both types) (50 verbs), like *encarcelar* 'jail' and *ensillar* 'saddle'.
- Other change of state and change of location verbs (50 vebs), like *romper* 'break', *venir* 'come'.

Some of these verbs are also further classified based on their morphology. As will be shown soon, this distinction will be useful to explore the idea that analogy may have played a relevant role in this development. Specifically, there are prefixed verbs among degree achievements and object experiencer psychological verbs. Locative verbs are always prefixed. For the sake of clarity and simplicity, I consider a verb derived only if it transparently shows the prefixes *en-* and *a-* and a noun or adjective. For example, *alargar* 'lenghten' and *asustar* 'frighten' are derived, as they are formed by the prefix *a-* plus the adjective *largo* 'long' or noun *susto* 'fright'. Verbs like *amargar* 'get bitter' or *amar* 'love', however, are considered morphologically simple, as it is not clear that they are prefixed due to the fact that their related adjective and noun (*amargo* and *amor*, respectively) begins with the same vocalic sound. Locative verbs like *encarcelar* 'jail' and *engrasar* 'grease' are derived, as it is clear that these verbs have the prefix *en-* and have related nouns *cárcel* 'jail' and *grasa* 'grease', respectively.

5.4.3 Results

In this section I report the results of exploring the development of *ser* and *estar* in combination with the verb classes discussed in the previous section. Figure 5.3 illustrates the changes in the relative frequency of *estar* with each class of verb, as opposed to *ser* with the same types of predicates. The frequencies here are relative to the total number of both *estar* and *ser* with each of these classes. Tables 5.7 and 5.8 show the total numbers and frequencies in percentage points. Appendix

E contains tables indicating the total number of *estar* and *ser* with these types of verbs.

As can be clearly seen from the graph 5.3, the relative frequency of *estar* increases considerably with all verb classes from the 14th century. From the graph, it is also clear that there are two clear stages in the development of *estar* + PTCP (as opposed to *ser* + PTCP):

A. Period of complex or inconsistent variation from the 13th to the 15th century, when the frequency of *estar* with all classes of verbs as opposed to *ser* is low.

B. Period of extension from the 15th to the 20th century. After the 15th century, change speeds up first with locatives, extent and object experiencer psych predicates and other change of state verbs. The frequency of *estar* with the other classes of verbs increases slightly until the 17th or 18th century, but then it either drops (with incremental theme and change of location verbs) or stabilizes (with locatives and degree achievements) to yield a similar pattern of use.

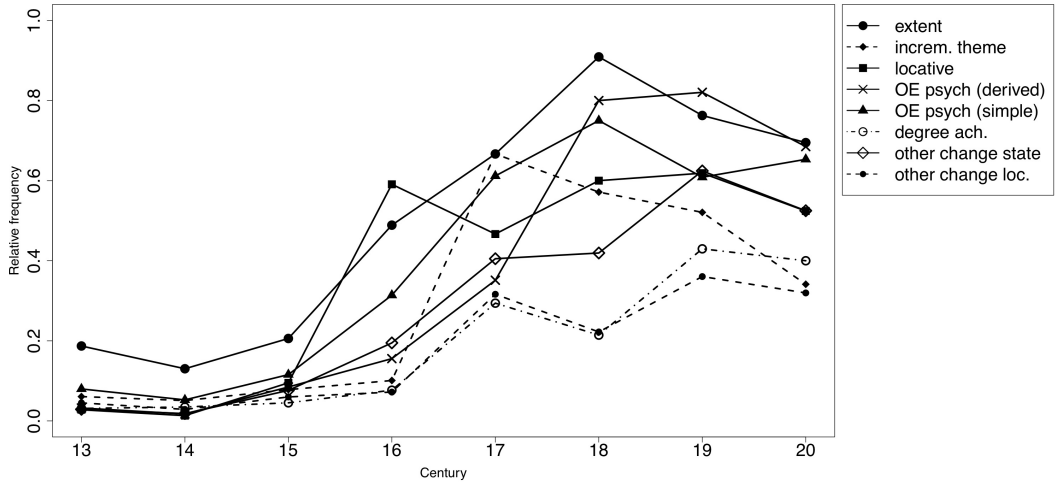

Figure 5.3: Frequency of *estar* with extent predicates, incremental theme verbs, locative verbs, morphologically simple and derived object experiencer verbs, degree achievements, and other change of state and change of location verbs.

Interestingly, extent predicates are the class of verbs which appears more frequently with *estar* since the earliest centuries, reaching almost 20%. The other

classes of predicates are combined with *estar* in less than 10 per cent of the cases, with a slightly higher frequency only in the case of object experiencer psych verbs.

Furthermore, the increase in the frequency of *estar* with object experiencer psych verbs parallels the increase with extent predicates to converge in the 20th century. Locatives and change of state verbs also converge in their frequencies with *estar*, but after very dramatic changes with locatives and, quite differently, a very steady increase with change of state verbs from the 15th to the 19th centuries. On the other hand, the spread of *estar* through degree achievements and change of location or motion verbs is almost identical. It is also interesting to note that the frequency of *estar* with incremental theme predicates, however, after reaching its highest peak in the 17th century, decreases in frequency, converging with degree achievements and other change of location verbs in the 20th century.

In sum, in the latest centuries, when *estar* + PTCP is the only expression for adjectival passives, *estar* has already outnumbered *ser* with extent predicates, object experiencer psych verbs, locative and other change of state verbs, whereas it remains to be outnumbered by *ser* with degree achievements, incremental theme and change of location or motion predicates.

The corpus evidence, as just presented, suggests that the contexts of principal relevance for the spread of *estar* as the only expression for adjectival passives are those, on the one hand, in which the copula is combined with object experiencer psychological predicates and other extent predicates and object experiencer psych verbs, and, on the other, with locative verbs and other change of state predicates. How can this extension of adjectival passives with *estar* be explained? In what follows I try to begin to answer this question in terms of analogy and priming.

	Extent		OE psych		Locative		Change of state	
	%estar	Total	*%estar*	Total	*%estar*	Total	*%estar*	Total
13th c.	18.7	668	7.7	888	2.8	358	2.9	4425
14th c.	13	307	4.7	724	1.3	455	1.8	3383
15th c.	20.6	544	10.7	1082	9.5	284	7.6	3097
16th c.	48.9	270	30	510	59.1	66	19.5	1247
17th c.	66.7	30	53.3	122	46.7	15	40.5	158
18th c.	90.9	11	76	25	60	5	41.9	31
19th c.	76.3	274	68.5	1341	61.9	139	62.4	764
20th c.	69.5	141	66.2	476	52.4	124	52.5	381

Table 5.7: Percentage of verb classes with *estar* (from the total number of *ser* and *estar* with these verbs) from the 13th to the 20th century.

	Degree achiev.		Increm. Theme		Change of location	
	%estar	Total	*%estar*	Total	*%estar*	Total
13th c.	3.1	897	6.1	1267	4.5	3548
14th c.	3.5	489	5.1	609	2.9	2468
15th c.	4.5	1018	7.8	832	6	2571
16th c.	7.7	1003	10.1	624	7.2	1722
17th c.	29.4	34	66.7	51	31.6	98
18th c.	21.4	14	57.1	14	22.2	18
19th c.	43	242	52.1	288	36	405
20th c.	40	140	34.1	126	32	219

Table 5.8: Percentage of verb classes with *estar* (from the total number of *ser* and *estar* with these verbs) from the 13th to the 20th century.

5.4.4 Discussion

The development of *estar* with the types of verbs described above, as just described, can be better understood if we take the general cognitive mechanism of analogy into account. The first contexts of use of *estar* with participles are those in which base predicates hold some (morphological and semantic) similarities with prepositional phrases appearing with *estar* in the earliest centuries. Analogy or pattern matching (both at the semantic and morphosyntactic level) is the mechanism that can explain this change.

To argue for this idea, we need to establish which were contexts of use of *estar* in the earliest centuries. In Old Spanish, *estar* could be used as a postural verb, as in (12). Soon afterwards, however, these uses almost totally disappeared.

(12) *A todos los sos* ***estar*** *los mando*
to all the his be.LOC them order
"And he orders that all his [knights] should stay there" (12th c., CID)

At that time it could also be used, as in Late Latin, with prepositional phrases describing the spatial location of an entity, coded as the subject, as in (13).

(13) ***Estando*** ***en la cruz*** *vertud fezist muy grant*
being.LOC in the cross virtue made very great
"By staying in the cross you made an act of goodness" (12th c., CID)

Since the earliest centuries *estar* could also combine with prepositional phrases describing a more figurative or abstract location. For example, in (14-c), where the prepositional phrase refers to a psychological state, or in (14-a), in which it refers to a nominalized process. It could also refer to a previous state of affairs, as

in (14-b), with a neutral pronoun.

(14) a. *Yo lo veo que* ***estades*** *uos* ***en yda***
I it see that are.LOC you in gone
"I see that you are on a journey" (12th c., CID)

b. *Ellos* ***enesto estando*** *don auien grant pesar*
they in this being.LOC where had great sorrow
"In this (being), where they had great sorrow..." (12th c., CID)

c. *el rey no paraua mientes sino a comer & beuer: &*
the king not stopped minds unless to eat and drink and
estar *toda via* ***en vicio*** *& mas señaladamente con mugeres.*
be.LOC still in vice and more especially with women
"The king did not stop eating and drinking and to giving himself over to vice, especially with women" (13th c., VLT)

These contexts of use of *estar* and the loss of the postural uses of this verb suggest that *estar* was undergoing some change during these centuries. A change from a postural to an underspecified meaning to *estar* could explain this wider range of contexts in which the verb could be used, always in combination with prepositional phrases. After this change happened we expect that *estar* could combine with a wider range of complements like participles as well as with other morphological categories like adjectives or gerunds. On the other hand, if there is some similarity relation established between the original contexts of use of *estar* and the first participles combining with *estar*, then we would expect that the first participles appearing with *estar* are those which describe some locative relation, such as extent predicates, as in (15), or locative verbs, as in (16).

(15) a. *Ca estaua ya açerca; del otro emperador. sso*
because was.LOC already close of the other emperor under
que ***staua*** ***cercado*** *de gente a derredor*
that was.LOC surrounded of people to around
"Because he was already close to the other emperor, and he was surrounded by people" (13th c., ALX)

b. *E por que esta villa* ***estouo*** ***çercada*** *muchos dias*
and by that this village was.LOC surrounded many days
entre tanto acaesçio enel Reyno algunas cosas que son
between so many happened in the kingdom some things that are
a contar
to tell
"And as this village was surrounded for so many days, during that time there happened in the kingdom some things which are worth

telling" (13th c., CAX)

(16) a. *Los peones [...] que* ***estan*** ***armados*** *&* *guisados*
the workers that are.LOC armed and prepared when
quando quier lidiar.
want fight
"The workers, who are armed and prepared when they must fight" (13th c., ACE)

b. *E agora ya es el pueblo muerto. & el templo arde pues*
and now already is the people dead and the temple burns as
uos porque ***estades*** ***armados****. dexat las armas.*
you because are.LOC armed leave.IMP the weapons
"And now the people are dead, and the temple burns. And you, why are you armed? Leave the weapons" (13th c., EE1)

Intuitively, underlying these predicates there is some locative description which makes them similar to locative prepositional phrases like *en la cruz* 'in the cross'. For example, *cercar* 'surround', which appears 91 times with *estar* in the 13th century, describes a locative relation between a figure and a place or location, encoded as the grammatical subject and object respectively, as in (17).

(17) a. ***Çercar*** *quiere a valençia pora xristianos la dar*
surround want to Valencia for christians it give
"He wants to surround Valencia, so that he can give it to the Christians" (12th c., CID)

b. *Dos mato con lança & .v. con el espada Los moros son*
two killed with spear and 5 with the sword the moors are
muchos derredor le ***çercauan***
many around him surrounded
"He killed two with a spear, and 5 with the sword. There were many moors surrounding him" (12th c., CID)

The predicate *armar*, which appears 90 times with *estar* in the 13th century, describes a different type of locative relation between its participants. Specifically, its meaning can be paraphrased as 'put X on/in Y', where X is the noun which is transparently seen in the verb. In the previous example the noun is *arma* 'weapon'. that the verb takes its name from and Y is the figure, encoded as the internal argument. These are so-called **locatum** verbs in the literature (Clark and Clark, 1979; Levin, 1993; Buck, 1993). Similar English verbs are, for example, *saddle* as in 'saddle a horse' which means 'to put a saddle on a horse'. This is illustrated in sentences (18):

(18) *E otrosi los arrayazes dela mar dixeronle que mandase* ***armar***
and then the captains of the sea told him that sent arm
aquellas catorze galeas de buenas gentes & bien armadas
those fourteen ships of good people and well armed
"And then the captains of the sea told him to arm those 14 ships with good and well armed people" (13th c., CAX)

The idea that there is some semantic similarity between the prepositional phrases that could appear with *estar* in the earliest centuries, as described before, and the first participles combining with *estar* is not new. It was already suggested by Bouzet (1953), who does not elaborate on it further. This idea goes as follows. In analogy an already existing relation becomes generalized to other linguistic forms; by this mechanism new forms are created or assimilated on the basis of existing patterns. A linguistic item or construction A is generalized to other linguistic forms by aligning the new linguistic form B with the existing linguistic form A, which serves as a model. Specifically, analogies are constrained on the basis of the similarities arising between the components of the source and target domains. After this, the new analogical relation can be applied to the target domain, which eventually leads to a restructuring or reanalysis of such domain. In this view, a new linguistic form or expression is not created from scratch, but is based on an existing one, which acts like a pattern. Eventually if an innovative form created on the basis of such analogical relation succeeds, it spreads through the community of speakers in the language. I propose that such changes are built from complex analogical relations made by speakers on the basis of both morphology and semantics. The spread in language change occurs through the repetition step, which can be modelled through frequency of use of the reanalysed target domain.

Going back to our case study, the analogical relation that may explain the emergence and lexical extension of adjectival passives with *estar* + PTCP was two-sided: on the one hand, those cases of *estar* with extent and locative predicates emerged by similarity with prepositional phrases describing the spatial location of an entity; on the other, by analogy with more abstract or figurative locative prepositional phrases *estar* begins to combine with psychological predicates and change of state verbs.

The innovative form *estar* + PTCP was created precisely in this way, as the first participles combining with *estar* are chosen on the basis of their similarity with locative prepositional phrases. However, there is a crucial difference between the traditional view of analogy and the account proposed here. In the present account, not only morphology but also semantics helps to establish these relations; specifically in this case the locative meaning of prepositional phrases appearing with *estar* influenced the type of predicates combining first with this

copula. There is evidence from language acquisition studies that analogies, both at the morphosyntactic and semantic level, plays an important role in language learning as well (Goldwater *et al.*, 2011).

Additional evidence for the idea that there exists a similarity (which make it possible to establish this analogical relation) between participles first combining with *estar* and locative prepositional phrases comes from examples such as those in (19), where the participle *encastillado* 'imprisoned, lit. in-castle-d' ((19-b)) is similar in its form to the locative prepositional phrase *en castillo* 'in a castle, lit. in castle'; or in (20), where participles are coordinated with locative prepositional phrases, as if they were the same category.

(19) a. *Ca el que es de buen coraçon sabe sofrir & lidia*
because the that is of good heart knows suffer and fights
esforçada mente commo sy ***estudiese*** ***en Castillo***
hardly as if were.LOC in castle
"Because the one who has a good heart knows how to suffer and fights as if he were in a castle" (13th c., CZP)

b. *Ffablades commo omne: que* ***esta*** ***encastellado*** *Mas sy prender*
talk like man that is.LOC in castle-d but if take
uos puedo: de fuera de sagrado Seades bien seguro: que seredes
you can of out of sacred be well safe that be
colgado
hung
"You talk like a man who is in a castle; but if I can take you, you can be certain that you will be hung" (13th c., BER)

(20) a. *de la otra [parte] cayen muchos muertos* ***en tierra*** *&* ***feridos***
of the other part fell many dead in ground and were
Ca bien semejaua que los vnos de los otros non
wounded because well seemed that the ones of the others
aujen piedad
not had pity
"Many dead men fell to the ground and wounded; as it looked like the ones did not pity the others" (13th c., CZP)

b. *E en todo el so regno las espannas* ***estauan*** ***en paz*** *&*
and in all the his kingdom the spains were.LOC in peace and
assessegadas *so el sennorio de los Romanos.*
calmed down under the lordship of the romans
"And all the lands of Spain were in peace and calmed down under the lordship of the Romans" (13th c., EE1)

The analogical relation established between locative prepositional phrases and the first participles combining with *estar* is also obvious in those cases where *en* appears before the participle, as in (21).

(21) a. *et dioles çient cauallos* ***en seellados*** *&* ***enfrenados***
and gave them hundred horses in saddled and in brakes
"And he gave to them one hundred horses saddled and with brakes" (13th c., EE2)

b. *todo lo que* ***es en parrieillado*** *en sarten apareiado sera del*
all it that is in grilled in grill prepared be of the
sacerdot que lo offrece si quier sea con olio si quier no.
priest that it offers if wants be with oil if wants not
"All that is grilled in a pan will be of the priest who is offering it, either with oil or without" (13th c., IJ8)

Interestingly, in example (21-b) the copula appearing with the participles is not *estar* but *ser*, whereas this copula can never appear in this context in Modern Spanish. This further supports the idea that there was a competition between *estar* and *ser* to express adjectival passives.

There are other types of verbs appearing very frequently with *estar* in the earliest centuries which may apparently be more problematic for maintaining the analogy idea. These are cases of *estar* with object experiencer psychological verbs such as *espantar* 'frighten', which appears with *estar* 29 times in the 13th century, *asosegar* 'calm down' (27 times) or *airar* 'get angry' (11 times). The similarities of these verbs with locative prepositional phrases are not as evident as in the case of locative verbs. Some occurrences of *estar* with object experiencer psychological verbs from the corpus are given in (22).

(22) a. *Alexandre fue yrado. Manda lidiar cuemo* ***estaua***
alexandre was annoyed. order fight as was.LOC
ensannado.
in-rage-d
"Alexandre was annoyed. He made them fight because he was enraged." (13th c., ALX)

b. *ca* ***estauan*** ***espantados*** *de aquel golpe tan estraño.*
because were.LOC frightened of that knock so strange
"Because they were frightened by such a strange knock..." (13th c., CZP)

c. *Et ellos tanto* ***estauan*** ***enoiados*** *que lo touieron por bien.*
and they so were.LOC annoyed that it had for well
"And they were so annoyed that they understood it as good thing."

(13th c., EE2)

However, and despite appearances, object experiencer psychological verbs fit well into the syntax and semantics of *estar*, which in its origins was used to describe the position of an entity. Based on Gawron's (2005) analysis of the semantics of extent predicates and Sánchez-Marco and Marín's (2011) proposal that the same semantic component that exists in extent predicates is also part of the denotation of object experiencer psychological verbs, we assume that these verbs encode an abstract locative component (or *path*) in their meaning, which enables them to combine first with *estar*. The main idea, proposed in Sánchez-Marco and Marín, is that this type of verb requires an axis other than time to be anchored. The primary function of this axis is to locate and orient entities (experiencers or figures) on a scale in a particular domain, which is often provided by the noun related to the verb. This domain provides the anchor with respect to which the eventuality can be measured. Both the change of state meaning and stativity of these predicates can be explained if this path component exists in all three classes of predicates.

Examples suggesting the similarity of object experiencer psychological predicates with prepositional phrases in the earliest centuries comes from cases as in examples in (14-c) (repeated in (23) for convenience), where locative prepositional phrases describe a more abstract place or location; and other ocurrences as those in (24), in which the verb is introduced by *en*, similarly to examples with locative verbs, as in (21).

(23) *el rey no paraua mientes sino a comer & beuer: & **estar***
the king not stopped minds unless to eat and drink and be.LOC
*toda via **en vicio** & mas señaladamente con mugeres.*
still in vice and more especially with women
"The king did not stop eating and drinking and to giving himself over to vice, especially with women" (13th c., VLT)

(24) a. ***Era** esta mançeba de dios **en amorada***
was this girl of god in loved
"This girl was in love with God." (13th c., BER)
b. *njn pierdas verguença onde **finques en uergonçado***
nor loose shame where leave in shame
"Do not loose shame where you are ashamed" (13th c., CD1)

To sum up, we have seen how extent verbs, on the one hand, and on the other locative and psychological verbs, combine very frequently with *estar* since the earliest centuries. The usage frequency of *estar* with other classes of verbs increases, such as degree achievements and incremental theme verbs, increases later. I sug-

gest that priming may have been the mechanism which explains how adjectival passive *estar* + PTCP eventually spreads through the lexicon, in that the most frequent copula-verb combinations had a bigger impact in the development of the construction.

5.5 General discussion and concluding remarks

To sum up, change in Spanish participial constructions is a regularization change which leads to one interpretation for each periphrasis. On the basis of corpus evidence, we saw how one prevalent expression, say *ser* + PTCP, loses terrain to other form-meaning pairs. There is some competition between new (*estar* and *haber* periphrases) and older forms (*ser* + PTCP) describing similar interpretations which leads to an eventual replacement of one of the forms.

I have also explored in this chapter how change in participial constructions spreads through different classes of predicates, both in the case of the development of perfects and adjectival passives. In the case of the extension of adjectival passives with *estar*, we have seen how the original meaning of the copula influences the way the participial construction emerges and further develops, by a mechanism of analogy or pattern matching. In turn, the spread of the perfect with *haber* + PTCP suggests that this construction was highly grammaticalised as a perfect since the earliest centuries.

6

ANNOTATION AND REPRESENTATION OF THE CORPUS

6.1 Introduction

The diachronic corpus of Spanish used in this dissertation, whose contents are described in chapter 2, was automatically enriched with linguistic information of lemma and morphological class. In this chapter I present the method used in order to enrich the corpus with this information and to represent this information in a machine-readable format, so that it was possible to automatically obtain all combinations of auxiliary-participles from it.

As mentioned in chapter 2, only *Lexesp* texts already contain linguistic information of lemma and morphological class. Therefore, in order to carry out an empirical study of change, texts from the Middle, Modern and Contemporary periods needed to be enriched with this type of information. Unfortunately, when I started this project in 2009, there was no tool freely available that could directly be used to enrich Old Spanish texts with lemma and morphological class. Thus, motivated by this need, a secondary goal in this dissertation has been to develop a strategy to automatically enrich with linguistic information of lemma and morphosyntactic tags texts from the Middle, Modern and Contemporary periods. Basically, the strategy has consisted in adapting an existing linguistic analyzer for standard Modern Spanish in order to enable it to deal with Old Spanish.

The contents of this chapter are as follows. After introducing in Section §6.2 the main differences between Old and Modern Spanish, which make the strategy of directly using existing NLP tools for (standard) Modern Spanish insufficient, I move to Section §6.3, where I present the strategy used in this dissertation to

automatically annotate the corpus with linguistic information. Then, in §6.4, I present the results of a careful evaluation over automatically assigned lemmas and morphosyntactic tags in the corpus. I conclude this chapter in Section §6.5 with a description of the annotations used to represent information in each document about author, title, etc.

6.2 Differences between standard Modern and Old Spanish

Before presenting the strategy followed in order to enrich the texts with linguistic information, I will introduce some details about the differences between Modern and Old Spanish.

The apparently simplest and most basic strategy that we could use in order to automatically enrich Old Spanish texts with linguistic information is to directly use an existing NLP tool for Modern Spanish. There are, however, some notable differences between Old and (standard) Modern Spanish which make the apparently simpler strategy of directly using an existing NLP in order to enrich Old Spanish texts unsuccessful. The most obvious difference between Modern and Old Spanish is spelling variation. Spelling variation in Old Spanish texts is very noticeable throughout the medieval period. Several variants of the same word can be found not only in texts created within the same general historical period but also within the same text even when this has been transcribed by a single scribe. Obviously, old variants of words such as these are not contained in the dictionary of the standard Spanish analyzer. For example, the third person singular of the imperfective past of *haber*, that is *había* 'had', in Old Spanish had at least 20 spelling variants, as in (1):

(1) Spelling variants of *había* 'had' in the 13th century:
auie, hauia, auia, auja, auje, hauie, auya, avie, augua, avya, avia, auye, abia, hauja, avye, avja, hauya, abye, hauje, abya

A certain normalization in the spelling of the words can be seen in the texts produced by public notaries during the kingdom of Fernando III (1217-1252) and Alfonso X El Sabio (1252-1284), who followed the uses of the Castilian chancellery. However, it is not until the 15th century that a certain unification in the spelling uses is observed. In the normalization of the spelling the most noteworthy date is 1517, when the *Reglas de Ortographía* written by Nebrija were published (Nebrija, 1517). Two centuries later, after the establishment of the Real Academia Española,[1] a new era of normative grammars and spelling rules began, with the

[1] `http://www.rae.es`

subsequent decrease of spelling variants in texts. Different factors could play a role in the type of spelling variation present in the Old Spanish texts. Among them, the influence of Latin and paleographical and typographical factors stand out. As it is well-known, Latin was the most prestigious language throughout the Middle Ages in Western Europe. In the absence of clear spelling rules for the emerging Romance varieties, Latin was the only available model for scribes. Paleographical factors, such as the available space on the folio or the typography being used, could also influence the choice of one spelling variant (Torrens, 2002; Sánchez Prieto, 2005).

Although at first sight this variation seems arbitrary, some regularities can be found. For example, the *ñ* usually appears in old texts as *n* or *nn*, as the Latin geminated variant (*donna/dona*, 'mistress'). The characters *i*, *j*, and *y* are also found in the texts representing the same sound. The *u* represents both a consonant and a vocal sound. Thus, besides representing the vocalic sound */u/*, the letter *u* was also often used to represent a consonantic sound (possibly a voiced labiodental fricative or perhaps a voiced bilabial fricative as some authors have suggested) in words such as *cauallo* 'horse' or *ueer* 'to see'.

Other differences between Modern and Old Spanish concern syntax and punctuation or sentence breaks. Some syntactic features which are characteristic of Old Spanish are for instance some verb second phenomena and the position of clitics (Fontana, 1993). As for punctuation, it was also very different in Old Spanish. For example, there is a much more extensive use of colons and semi-colons in earlier historical periods, and sometimes there seems to be the case that ampersand symbol which stands for *y* 'and' also sometimes apparently is used to break sentences.

As an example from the 13th century, see (2), for which a possible "translation" into Modern Spanish is provided in (3):

(2) Llos meiores que pudo; en greçia escoger. Que lo sopiessen; en las .vij. artes enponer. Aprendia delas .vij. artes; cada dia liçion. de todas cada dia; fazie disputaçion

(3) Los mejores que pudo en Grecia escoger. Que lo supiesen en las .vij. artes imponer. Aprendía lección de las .vij. artes cada día. De todas cada día hacía disputación.

6.3 The method: Extending the tool

The method used in order to automatically enrich the corpus with linguistic information consists of adapting an existing linguistic analyzer for standard Modern Spanish in order to enable it to deal with the oldest texts in the corpus. Very

briefly, the method used to extend this tool consists basically of an expansion of the dictionary, a customization of other modules such as tokenization and affixation, and a retraining of the tagger.

The method and an evaluation of the results is presented in Sánchez-Marco *et al.* (2011). However, in the last year I have significantly enhanced the results obtained in the tagging by adding a notable amount of words to the dictionary as a well as a bigger training corpus. On the other hand, to annotate the corpus used in this thesis I have used the tagger available to the tool that combines constraint-grammar-like rules and probabilities. In Sánchez-Marco *et al.* (2011), in contrast, I reported the results obtained only using the probabilistic tagger included in the tool. Thus, in this section I present the strategy and results of a careful evaluation of this improved version, which has been used to annotate the corpus used in this dissertation, and compare it with the results obtained in Sánchez-Marco *et al.* (2011), which is already freely available in the version of Freeling 3.0 (Padró, 2011; Padró and Stanilovsky, 2012).

6.3.1 The analyzer: FreeLing

The linguistic analyzer used is FreeLing. FreeLing is a developer-oriented library providing a number of language analysis services, such as morphosyntactic tagging, sense annotation or dependency parsing (Padró *et al.*, 2010). This tool is open source, actively developed and maintained, and highly modular, and therefore it is particularly well suited for the purpose of tagging Old Spanish. In its current version (3.0), this resource provides services (to different extents) for the following standard language varieties: English, (standard) Modern Spanish, Portuguese, Italian, Galician, Catalan, Asturian, and Welsh. Besides, FreeLing 3.0 already includes the version of the analyzer for Old Spanish as it is presented in Sánchez-Marco *et al.* (2011).

The FreeLing processing pipeline for morphosyntactic tagging is illustrated in Figure 6.1. As shown in the figure, a set of texts is submitted to the analyzer, which processes and enriches the texts with linguistic information using different modules: tokenization, dictionary, affixation, probability assignment and unknown-word guesser[2], and PoS tagger.

The tagset used by this tool is based on the EAGLES standard[3]. The first letter of each tag indicates the morphological class of the word. The remaining letters

[2]This module has two functions: first, it assigns an *a priori* probability to each analysis of each word. Second, if a word has no analysis (none of the previously applied modules succeeded to analyze it), a statistical guesser is used to find out the most likely PoS tags, based on the word ending.

[3]Expert Advisory Group on Language Engineering Standards (`http://www.ilc.cnr.it/EAGLES96/home.html`).

(up to 6) specify more fine-grained morphosyntactic and semantic information, such as the gender and number of nouns or the tense, mode and type (main or auxiliary) of verbs. For example, the tag VMIP3S0 represents the third person singular of a main verb in the Indicative mood like *canta* 'he/she sings'.

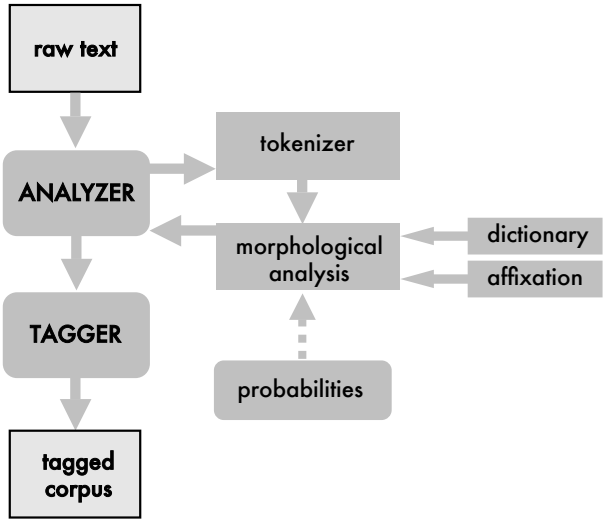

Figure 6.1: Processing pipeline in FreeLing.

6.3.2 Data

In order to adapt the tool, I have worked with the HSMS texts, which contain a representative sample of the genres for each century from the 12th to the 16th century, as described in chapter 2. The differences between standard Modern and Old Spanish since the 17th century are significantly less and therefore this part of the corpus was not used to adapt the tool.

On the other hand, a *Gold Standard Corpus* has been created in order to retrain the tagger and to carry out the evaluation and the error analysis. This corpus contains approximately 60,000 words, half of which contains a sample of texts from the HSMS texts (as described in Sánchez-Marco *et al.* (2011)) and the other half contains a sample of texts from *Lexesp* corpus. Thus, this Gold Standard Corpus contains 30,000 more than in the version presented in Sánchez-Marco *et al.* (2011). The part of this corpus containing a sample of the HSMS texts has been pre-annotated using the Standard Spanish tagger and the Old Spanish dictionary and, then, manually corrected. In this tagging the expanded dictionary was used, which has minimised the time needed to correct the corpus considerably.

The effort of correcting the corpus is much lower compared to annotating from scratch.

6.3.3 Dictionary expansion

Data. The Modern Spanish dictionary contains 556,210 words (669,121 lemma-tag pairs). This dictionary has been expanded with 58,435 new word forms or entries, totalling 744,160 lemma-tag pairs, thus increasing the number of word forms in the dictionary to more than 614,000. For example, the word form *y* in the expanded dictionary has four different lemma-tag pairs, corresponding to a coordinate conjunction, a noun, a pronoun, and an adverb, whereas in the Standard Spanish dictionary it has only two lemma-tag pairs, corresponding to the coordinate conjunction and noun uses.

Table 6.1 illustrates the distribution of the categories of words which have been added to the dictionary. As could be expected from the general distribution of words across PoS categories, verbs and nouns account for more than half of the words added. In contrast with the standard Spanish dictionary, in the expanded dictionary I also added a list of proper names and some Roman numbers. The reason to do this is that in Old Spanish both category are often written in lowercase letters and, in the case of Roman numbers, they show as well inconsistent annotation marks or even they do not follow standard rules of composition.

Apart from the considerably greater number of words added to the dictionary in comparison with the version in Sánchez-Marco *et al.* (2011) (specifically 26,420 words more have been added), there is also a difference regarding the distribution of the words added. In the version presented in this thesis the percentage of verbs and nouns is significantly bigger (almost 50% more for verbs and seven percentage points more for nouns). This difference is partly due to the fact that most participles have been manually corrected and added to the dictionary as new entries as well. This step has been done in order to increase the precision and the coverage in the retrieval of participial constructions, which is the case study in this dissertation.

Verbs	83.4%	Pronouns	1.3%
Nouns	26.8%	Determiners	1%
Adjectives	9.4%	Adverbs	0.7%
Prepositions	2.1%	Conjunctions	0.5%
Numbers	1.7%	Interjections	0.3%
Proper names	1.4%	Punctuation	0.01%

Table 6.1: Distribution of words added to the dictionary.

Another difference between the dictionary in this version and the version presented in Sánchez-Marco *et al.* (2011) is the way joined words such as *del* 'of the' are represented. Specifically, in the previous version joined words are stored in the dictionary as a distinct word with one lemma. For example, the corresponding lemma to the word *del* 'of the' is the same form *del* and the tag is *SPCMS*, where the letter *C* indicates that the word is joined. The problem with this approach is that the information about the original composition of such word, that is preposition *de* 'of' plus determiner *el* 'the' in the example, is lost. To solve this problem, the two distinct lemmas composing joined words have been stored in each entry. For example the entry for the word *del* is *del del_de+del_el SP+DA*, where both the information about the preposition and the determiner in the masculine singular form can be retrieved from the dictionary in the tagging process.

Method. I used two different strategies to automatically generate the entries to be added to the dictionary. The first strategy consists in using two types of mapping rules: substring rules and word rules, as presented in Sánchez-Marco *et al.* (2011).

Substring rules map 54 sequences of characters from an old variant onto the corresponding standard variant. These mapping rules are based on the observed regularities in the spelling of Old Spanish texts (Sánchez Prieto, 2005; Sánchez-Marco *et al.*, 2010). These rules are independent of the morphophonological context, except that 18% of them are restricted to the beginning or the end of a word. Table 6.2 shows some examples of these rules. 42% of the types added to the dictionary have been generated using these rules. All words generated by this method are added to the dictionary if and only if they are contained in the corpus. This avoids the automatic generation of a very high number of variants.

Old	Modern	Example
euo	*evo*	*nueuo* → *nuevo* 'new'
uio	*vio*	*uio* → *vio* 'saw'

Table 6.2: Examples of the substring rules.

39% of the types incorporated into the dictionary have been created using **word rules**. These are mappings from an old variant of a word to its corresponding standard variant (created manually), to deal with the most frequent types not covered by the substring rules, such as for instance words without an accent (*consul* → *cónsul* 'consul'), or other graphemic variants (*yglesia* → *iglesia* 'church', *catholica* → *católica* 'catholic').

In order to assign which lemma corresponds to each word form, the following criteria have been followed. If the old variant has a graphically similar word which

is in a Modern Spanish dictionary (in this case, I used the *Diccionario of the Real Academia Española*,[4]) then it is assigned the corresponding standard lemma. A word is considered very similar if and only if the old variant can be obtained with a few number of transformations. If the old word does not have a very similar variant, the old variant is kept. For example, words only existing in Old Spanish like *maguer* are represented in the dictionary with the same old variant.In other cases, for example when there is no similar standard variant but the word shows an inflected form, then lemma of the corresponding word was reconstructed with the minimum modification. For example, the participle *valentiniado* in masculine singular form has been assigned the lemma *valentiniar*.

The remaining 19% of the added dictionary entries have been automatically created using **VARD 2**, which is a tool intended to be a pre-processor to other corpus to assist with spelling variation especially with Early Modern English texts (Rayson *et al.*, 2007; Baron and Rayson, 2008; Baron, 2011). This strategy is new and it has still not been used, to my knowledge, to expand Old Spanish dictionaries. It was also not used in Sánchez-Marco *et al.* (2011). VARD 2 uses techniques derived from modern spell checkers to find candidate modern form replacements for spelling variants found within historical texts. See more detailed information in the references mentioned above and in the webpage of the tool.[5] This tool has been used here to automatically generate variants for Old Spanish words, which were not previously covered by the first method. These generated variants have been used as a mapping from the old variant to a word in the Old Spanish dictionary, whose lemma and PoS tag was automatically assigned to the old variant and incorporated then to the expanded version of the dictionary.

In order to use this tool, a list of 69 spelling rules was ellaborated on the basis of the ortographic differences between (standard) Modern and Old Spanish. These are more general rules that the previous substring mapping rules used in the expansion of the dictionary. As can be seen from examples of these rules in Table 6.3, instead of mapping two or three characters, as it is the case in substring rules, they are rules of one, or two characters at most. Another difference with respect to the substring rules is that here mappings into accentuated vowels are included as well, so that a large number of not-accentuated words in Old Spanish was covered in the tagged version of the corpus. Of these rules only 9% are restricted to the beginning or end of word and the remaining 91% can be located in any part of the word. After several trials, the better results with this tool were obtained using a Threshold 50% and a Recall 1/4.

[4] `http://www.rae.es`.

[5] `http://www.comp.lancs.ac.uk/~barona/vard2/`

Old variant	New variant	Context
j	*í*	anywhere
nn	*ñ*	anywhere
rr	*r*	anywhere
sp	*esp*	start
nt	*nte*	end

Table 6.3: Examples of the spelling rules used in VARD 2.

6.3.4 Other modules

FreeLing analyzes forms not found in the dictionary through an **affixation** module that checks whether they are derived forms, such as adverbs ending in *-mente* or clitic pronouns (*-lo*, *-la*) attached to verbs. This module has also been adapted, incorporating Old Spanish clitics (*-gela*, *-li*) and other variants of derivation affixes (adverbs in *-mientre* or *-mjentre*). The lemma in these adverbials has not been normalized, but kept the same as in the original word, for example the entry for *firmemjente* is *firmemjente firmemjente RG* .

6.3.5 Retraining of the tagger

FreeLing includes two different modules able to perform PoS tagging: a hybrid tagger (*relax*), integrating statistical and hand-coded grammatical rules, and a Hidden Markov Model tagger (*hmm*), which is a classical trigram markovian tagger, based on TnT (Brants, 2000).

In this chapter I present the evaluation of the performance of the extended resource using both the *relax* and the *hmm* tagger with the probabilities generated automatically from the trigrams in the Gold Standard Corpus. As shown below, the accuracy in the tagging is significantly better with the relax than with the hmm tagger, and therefore the former has been used to annotate the corpus used in this dissertation.

The grammatical rules in the relax tagger used in this dissertation includes the rules in the Modern Spanish tagger plus a set of rules which I ellaborated in order deal with some potential ambiguities observed in Old Spanish texts. For example, *y*, which is a coordinate conjunction in Modern Spanish, by contrast in Old Spanish was also very frequently used as a deictic pronoun similarly to English *there*, as in *si alguna y falle* 'if anybody fails there'. By observing occurrences of this in the corpus, I ellaborated a rule to tag *y* as an adverb when it appears before a finite verb and after a noun, adverb, adjective, complementizer, or pronoun. Another rule which has been added to the grammar is that a verb should be tagged with the

grammatical information of first person singular if some positions before or after it there is a first person singular pronoun, e.g. *yo* 'I'. This rule has been added in order to solve ambiguities due to the homophony that exists in Old Spanish between first and third singular person verb forms, for example *canto*, which could mean 'I sing' or 'he/she sang'.

6.4 Evaluation

In this part I report the overall tagging results. The results for Standard Spanish have been used as a baseline for a state-of-the-art tagger for standard Modern Spanish (SS). I will show how the results have been significantly improved in this version (referred to as 3.1) by comparing it with the results of the tagging as presented in Sánchez-Marco *et al.* (2011) (3.0).

6.4.1 Dictionary

In order to evaluate the expanded dictionary, we use three different measures: ambiguity, coverage, and accuracy of entries generated using VARD 2:

- **Ambiguity** measures the average number of lemma-tag pairs per word form. To compute average ambiguity, each word form is assigned a score corresponding to the number of lemma-tag pairs in its dictionary entry. We have checked ambiguity in two different ways: (i) in the dictionary (type-based), (ii) in the corpus (token-based).

- **Coverage** measures the percentage of tokens in the corpus which are analysed by the dictionary. Uncovered or unknown words are those forms which are not included in the dictionary or analysed by the affixation module.

- I also evaluated the **precision** of automatically generated entries, that is the percentage of correct words among those added to the dictionary using VARD 2, and the percentage of the expected lemmas for those words actually added by the rules.

Both precision and recall have been obtained of the version used this dissertation by checking a random sample of 218 types (corresponding to 2% of the types created using VARD 2).

The results of the evaluation are summarised in Tables 6.4, 6.5 and 6.6. As can be seen from Table 6.4, ambiguity obtained in the version prepared for this thesis is similar to the version in Sánchez-Marco *et al.* (2011). Even if almost the double number of entries have been added to version 3.1 than version 3.0, the

ambiguity obtained in the evaluation is similar. The Old Spanish Corpus is more ambiguous than the Standard Spanish Corpus, despite the fact that the dictionary is not. The higher ambiguity of both Old Spanish dictionaries is probably due to the fact that function words have more entries in the Old Spanish dictionaries than in SS dictionary and by the fact that the dictionary in Old Spanish contains words from a much larger time span.

In contrast, both the results obtained for the coverage and precision are different. Table 6.5 shows that the coverage obtained in the latest version 3.1 is almost two percentage points higher than in version 3.0, which indicates a significant improvement. In the tagging of the corpus words that are not covered (five in 100 in version 3.1) must be guessed by the probability module, which increases the number of potential errors. The improvement obtained here is also signficant, as in this new version there are three in 100 words less to guess than in version 3.0. The higher coverage is most likely due to the fac that many more words have been added to the dictionary in version 3.1 than in version 3.0.

Finally, Table 6.6 summarises the precision obtained in the entries automatically generated using substring rules and the one obtained using VARD 2. As can be seen from the table, precision obtained using VARD 2 is 5 points lower than the one obtained using the substring rules (as reported already in Sánchez-Marco *et al.* (2011)). This is probably due to the fact that substring rules are more context restricted than VARD 2 generated variants. Incorrectly generated entries are mostly cases of proper nouns *amin*, *noel* or *belio*, and Latin words e.g. *anatis*; other cases are graphically similar words and possible variants but which they simply does not exist, e.g. *conoxen* is from the verb *conocer* 'know' but it has been incorrectly associated to *convenir*, or *quesy* to *casi* instead of joined word *que* plus *si*.

	Type-based	Token-based
SS	1.20	1.68
3.0	1.21	1.85
3.1	1.21	1.84

Table 6.4: Token and type-based ambiguity of the dictionary obtained for version 3.0 and 3.1 and for SS.

6.4.2 Tagging

In order to evaluate the performance of the tagger, the *accuracy* in the tagging of lemmas, PoS-1 (the whole label, containing detailed morphosyntactic information; 6 characters of the tag in total), and PoS-2 (word class; 1 character in total)

	Token-based
SS	99.4%
3.0	92.6%
3.1	94.2%

Table 6.5: Coverage of the dictionary over the corpus obtained for version 3.0 and 3.1 and for SS.

	Lemmas	PoS
substring rules	99.2%	99.2%
VARD 2	94.5%	95.4%

Table 6.6: Precision of the entries in the dictionary obtained using substring rules and VARD 2

has been checked. In all cases, this measure has been obtained as a result of a 5-fold cross-validation over the Gold Standard Corpus. As described before, the method proposed involves (a) adapting the dictionary and other modules, (b) retraining the tagger with Old Spanish texts. To assess the relative impact of these two adaptations, in Sánchez-Marco *et al.* (2011) the results of evaluating the tagging under several conditions was reported. Here I will report the scores obtained using the (C0) original tools for Standard Spanish, the tools in version 3.0 (C1-hmm), and in the version 3.1.In order to assess the impact of the type of tagger used (hmm or relax), I also report here the evaluation over both taggers using the 60,000 Gold Standard Corpus (C2-hmm) and the relax tagger (C2-relax), which is the version used to tag the corpus in this dissertation.

The accuracy scores obtained on the Gold Standard Corpus are summarised in Table 6.7. This table shows that in each of the conditions, the accuracy increases. Most of the improvements are significant at a 99% confidence level (χ^2 test, 1 d.f.). As can be seen from row C0, the results of directly applying the existing Modern Spanish analyzer over Old Spanish texts is around 70%, which is not fatal, but still insufficient to conduct research in historical linguistics precisely. Crucially, I have obtained an improvement of more than three points for lemma and morphological class and almost two points in the whole tag in comparison with the version presented in the Sánchez-Marco *et al.* (2011) (C1-hmm). Under conditions C2, there is a significant improvement in the PoS-1 using the relax tagger, whereas no significant improvement is obtained for PoS-2 and lemmas.

The results indicate that both adapting the dictionary and other modules and retraining the tagger have a positive impact on the overall performance of the extended tool on Old Spanish texts. The best result with the full set of tags (PoS-1)

	Lemma	PoS-1	PoS-2
C0	72.4	70.9	77.4
C1-hmm	92.6	89.9	94.5
C2-hmm	95.8	90.1	95.3
C2-relax	95.8	92.6	95.7
SS	99.1	94	97.6

Table 6.7: Accuracy obtained for lemma, PoS-1, and PoS-2 in the 5-fold cross-validation for the Old Spanish tagger on the Gold Standard Corpus (rows C0 to C2) and for Standard Spanish (row SS).

is 92.6% and 95.7% for the main PoS, which is a highly significant improvement with respect to the results reported in Sánchez-Marco *et al.* (2011) (C1-hmm).

To compare the Old Spanish and Standard Spanish taggers on the same basis, we retrained the FreeLing Standard Spanish tagger on a 30,000-token fragment of the LexEsp corpus. As can be seen from the Table in row SS, the results obtained for standard Spanish are still significantly higher for lemma (χ^2 test, 1 d.f., 99% conf. level, $p < .01$) and PoS-1 (χ^2 test, 1 d.f., 95% conf. level, $p < .05$) than those for the Old Spanish tagger: The accuracy over PoS-2 is 97.6%, almost 2 points higher than the 95.7% obtained for Old Spanish, and over lemma is 3 points higher. However, the results for the whole tag (PoS-1) are not significantly different (only 1 point of difference!), which indicates that in this version the results obtained in the tagger for Old Spanish are already very close to state-of-the-art taggers for standard language varieties. This is already a very significant improvement from the version presented in Sánchez-Marco *et al.* (2011). The error analysis presented below shows the causes of these errors, giving clues as to how this performance could be improved in future work.

6.4.3 Error analysis

The analysis of errors has been conducted over the 100 most frequent errors in tagging obtained with the Old Spanish tagger under condition C2-relax. This analysis shows that most of the errors in the tagging are due to the ambiguity in the dictionary, as could be expected given the discussion in the section 6.4.1. Specifically, 81.6% of the errors corresponds to words for which the correct tag is available in the dictionary, but the tagger has not selected it. Importantly, this percentage is more than 10 points lower than the one reported in Sánchez-Marco *et al.* (2011), which likely suggests that the relax tagger helps to deal with ambiguity better than the hmm tagger. More than half of these errors (73%) is due to types which are also ambiguous in the Modern Spanish dictionary. The most frequent errors in-

volve (i) function words such as determiner vs. clitic readings of *lo* 'the' or 'it' and relative pronoun vs. subordinating conjunction readings of *que* 'that'. The remaining 27% of the errors due to ambiguity are mostly words lacking the accent in Old Spanish. These are ambiguous verbal forms of the present and simple past (*llego* 'arrive' or 'arrived'), pronouns (*el* 'the|he'), and adverbs (*mas* 'more|but'). The remaining 18.4% of the errors that are not due to ambiguity correspond to words which were not added by any of the methods used to expand the dictionary, mostly proper nouns (*pierres, antolinez*), but also other words not covered by any rule (*fasta* 'until').

Now that we have seen the strategy followed to enrich the corpus with linguistic information of lemma and morphological tag, I will move to present how this information was represented in the corpus to that all occurrences of participial constructions and any other contextual information could be automatically retrieved from the corpus.

6.5 Representation

Prior to the representation of linguistic and structural information, all mark-up from the original texts was removed. Specifically, paleographic annotations present in the *HSMS* texts (which mark headers, hyphenation and other physical features present in manuscripts) as well HTML markup was removed from all texts coming from the *Gutenberg project* and *Biblioteca Cervantes*. (4) is an excerpt of the original edition of *Cantar de Mio Çid*, as included in the *HSMS* texts, where characters included between '<' and '>' indicate the expansion of an abbreviation by the scribe, or 'CB1.', which indicates that the following text is the first column in the given folio.

(4) {CB1.
Ca delo q<ue> mas amaua yal viene el mandado
Dozi[en]tos cauall<er>os mando exir p<r><<i>>uado
Q<ue> Rec'iban a myanaya & alas duenas fijas dalgo

Preserving paleographic information has been the focus of many philological research and is also relevant to consider in language change studies, especially in those cases where a difference between two copies of the same document may show significant divergences in a given phenomenon. However, for reasons of time and space all this information was simply removed from the texts using substitution scripts. An alternative proposal to preserve this information using standoff annotation is presented in Sánchez-Marco *et al.* (to appear).

The information encoded in the version used in this thesis contains both linguistic information of lemma and morphosyntactic tag, added using the extended version of the FreeLing analyzer, as presented before, as well as structural information about author and title, date of composition and century, genre or type of text, collection and identification code (ID), included in very simple xml-format tags (as tabulated in Appendix A). Then, the IMS Open Corpus Workbench[6] has been used to index the corpus and extract all occurrences and frequency counts of participial constructions.

An information which is crucial for any study on language change is the date of the documents composing the corpus. Such date can refer either to the year of composition of the text, or to a century or period. As it is well-known, deciding which is the date of an old document is not a trivial task, especially in those texts from the Middle Ages, which often have multiple copies and editions some times even from different centuries. In order to date each document in the corpus, such information has been taken from the sources of the documents. When the date is not available in the original source, other editions were checked. In these cases and for older documents, the date of composition has been chosen instead of the date of the copy of the manuscripts.

For texts coming from *Lexesp* the exact date of composition is not available. In the documentation dates of documents in the corpus are said to range from 1975 to 2000, but it is not specified which is the date of each document in particular (Sebastián-Gallés, 2000). In the quantitative study presented in the chapter 4, these documents have been randomly assigned a date ranging from 1975 to 2000. Using this strategy, a language change trend cannot be captured for this particular period, but at least we can still track the mean for this period.

6.6 Discussion

The results in the tagging over Old Spanish texts reported here are already close to state-of-the-art taggers for modern language varieties (almost 96% both over lemma and morphological category). These results are more than 20 percentage points higher than those obtained if the analyzer of Modern Spanish is directly used to tag Old Spanish texts. The corpus used in this dissertation to explore change in participial constructions has been annotated using the extended version of FreeLing, as described in this chapter. The results obtained in the evaluation guarantee a good retrieval of the occurrences of participial constructions (as presented in chapter 4), as well as the validity of the generalizations made on the basis of the quantitative analysis of data, as presented in Chapters 4 and 5.

[6] `http://cwb.sourceforge.net/`

In comparison to other state-of-the-art approaches the strategy proposed here is easily portable and reusable for other corpora and languages and yields a higher accuracy (Rayson *et al.*, 2007; Rögnvaldsson and Helgadóttir, 2008; Dipper, 2010, among others). Furthermore, the followed strategy builds on existing resources instead of building the whole resource from scratch, taking advantage of the existing similarities between standard and Old Spanish.

I leave for future versions of this corpus the annotation of punctuation and sentence breaking, for which a careful study of Old Spanish texts required, as well as the syntactic and/or semantic role labelling of the corpus, which may be useful for prospective studies.

7

CONCLUSIONS

This thesis is an empirical study of change in Spanish participial periphrases *estar* + PTCP, *haber* + PTCP, *ser* + PTCP, and *tener* + PTCP, from the 12th to the 20th century. In this thesis I have shown how the interpretations of these participial constructions have developed over time: From the complex variation system that existed in the earliest centuries, when each periphrasis could express more than two interpretations, to the Modern Spanish system, when each participial construction is the distinct morphological expression for only one interpretation. In Old Spanish the perfect could be expressed by all four periphrases, the stative possessive was instantiated by periphrases formed with *haber* or *tener*, and both verbal and adjectival passives could be expressed by *ser* + PTCP and *estar* + PTCP. In contrast, in Modern Spanish, perfects, stative possessives, verbal passives and adjectival passives, can only be expressed by *haber* + PTCP, *tener* + PTCP, *ser* + PTCP, and *estar* + PTCP, respectively.

I have argued, based on the qualitative and quantitative analysis of corpus evidence, that this development is an instance of regularization change, as complex variation has been eliminated from the participial construction system in the Spanish language. I have also proposed that this development may have been mainly motivated by the competition between these periphrases to express the perfect, adjectival and verbal passive and stative possessive interpretations, and explored the idea that the succeeding perfect and adjectival passive forms *haber* + PTCP and *estar* + PTCP spread through lexicon from some classes of verbs to others. In order to conduct this study I have compiled a large-scale diachronic corpus of Spanish, enriched with linguistic information of lemma and morphological tag using an adapted version of an existing linguistic analyzer (FreeLing).

The challenges that this dissertation has uncovered and that need to be tackled by future research are the following. To have a broader picture of the development of participial periphrases, it would be interesting to explore changes in participial constructions formed with other auxiliaries in the Spanish language, as well as to

conduct a crosslinguistic investigation of changes in participial constructions in other languages. In this respect, for example, it would be interesting to explore which type of variation show languages with auxiliary selection in the perfect, like German or French, at earlier centuries. Maybe more relevant for historical linguistics would be to investigate other regularization changes, as they happen throughout the centuries, and their relations with language acquisition studies about regularization. For example, it would be interesting to investigate whether children acquire auxiliaries with some semantic classes of verbs before than others. There are some studies about acquisition of copulas by children but they deal with the use of these copulas in combination with permanent-temporary properties (Schmitt *et al.*, 2004).

It would also be interesting to explore the development of these constructions using other (statistical) measures, such as for example distributional semantics measures (Schütze, 1998), or LNRE models of quantitative productivity Baayen (2001). Lastly, in order to further continue to develop resources to conduct empirical studies about language change in Spanish, it would be useful to enrich the diachronic corpus of Spanish used in this dissertation with syntactic as well as semantic information.

Appendix A

LISTS OF TEXTS IN THE CORPUS

The table in the following pages tabulates the list of all documents composing the diachronic corpus of Spanish used in this dissertation, ordered by date of composition. Columns in the table contain the following information:

1. Identification code (**ID**), or abbreviations used to refer to texts across the dissertation.
2. **Title** of the document. It is written *unknown* when the title is not mentioned in the sources.
3. **Author** of the document, or unknown when it is not clear in the source.
4. **Date** of composition of the document.
5. **Century**, given the date of composition of the document.
6. **Genre** or style of document.
7. **Collection**, or source of the document. The abbreviations used to refer to each collection in the corpus are as follows:
    - *ESMP* stands for *Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings* (O'Neill, 1999).
    - *PWAX* is *The Electronic Texts and Concordances of the Prose Works of Alfonso X, El Sabio* (Kasten *et al.*, 1997).
    - *FGNA* stands for *Textos y concordancias electronicos del Fuero general de Navarra* (Waltman, 1999).
    - MNAM is *The Electronic Texts and Concordances of Medieval Navarro-Aragonese Manuscripts* (Nitti and Kasten, 1997).
    - *TMEM* is *Textos medievales misceláneos* (Sanchez *et al.*, 2003).

- *HMSM* is *Concordancias electrónicas del corpus médico español* (Herrera and de Fauve, 1997).
- *ECTC* is *Early Celestina Electronic Texts and Concordances* (Corfis *et al.*, 1997).
- *CORDE* stands for new editions of texts from the electronic editions of the Hispanic Seminary of Medieval Studies provided by Sánchez-Prieto (p.c.). (Kasten *et al.*, 2011)
- *Gutenberg* stands for those texts coming from *Project Gutenberg* (`http://www.gutenberg.org/`),
- and *Cervantes* for those coming from *Biblioteca Virtual Miguel de Cervantes* (`http://www.cervantesvirtual.com/`).
- Lastly, *Lexesp* stands for *Lexico Informatizado Del Español* (Sebastián-Gallés, 2000).

ID	Title	Author	Date	Century	Genre	Collection
CID	Poema de Mio Cid	unknown	1100	12	poetry	ESMP
CZP	Libro del Cauallero Çifar	unknown	1214	13	prose	ESMP
SME	Santa María Egipciaca	unknown	1215	13	poetry	ESMP
VSM	Vida de Santa María Egipciaca	unknown	1215	13	prose	ESMP
BER	Obras de Gonzalo de Berceo	Gonzalo de Berceo	1230	13	poetry	ESMP
SMU	Semejanza del mundo	San Isidoro	1230	13	didactics	ESMP
TRO	Tres Reyes d'Orient	unknown	1230	13	prose	ESMP
FGN	Fuero general de Navarra	unknown	1237	13	law	MNAM
ALX	Libro de Alexandre	unknown	1240	13	prose	ESMP
FAR	Fueros de Aragon	Vidal de Canellas	1250	13	law	MNAM
ALB	Canones de Albateni	Alfonso X el Sabio	1250	13	didactics	PWAX
GE5	General Estoria V	Alfonso X el Sabio	1250	13	history	PWAX
JUZ	Judizios de las estrellas	Alfonso X el Sabio	1250	13	didactics	PWAX
LAP	Lapidario de Alfonso X	Alfonso X el Sabio	1250	13	didactics	PWAX
LEY	Libro de las leyes	Alfonso X el Sabio	1250	13	law	PWAX
MOA	Moamyn o Libro de las animalias	Alfonso X el Sabio	1250	13	didactics	PWAX
PIC	Picatrix	Alfonso X el Sabio	1250	13	didactics	PWAX
BDS	Bocados de oro	unknown	1250	13	didactics	ESMP
IJ8	Biblia I.I.8.	unknown	1250	13	religion	ESMP
PFG	Poema de Fernán González	unknown	1250	13	poetry	ESMP
POR	Poridat de poridades	unknown	1250	13	didactics	ESMP
REX	Dancus Rex	unknown	1250	13	didactics	ESMP
TEO	Doncella Teodor	unknown	1250	13	prose	ESMP
FRL	Fuero Real	Alfonso X el Sabio	1255	13	law	ESMP
CRZ	Libro de las cruzes	Alfonso X el Sabio	1259	13	didactics	PWAX
ZRQ	Tablas de Zarquiel	Alfonso X el Sabio	1260	13	didactics	PWAX
FJZ	Fuero Juzgo	Alfonso X el Sabio	1260	13	law	ESMP
LCB	Libro de los caballos	Teodorico Borgognoni	1260	13	didactics	ESMP
SPC	Espéculo	Alfonso X el Sabio	1260	13	law	ESMP
FN2	Fuero de Navarra	unknown	1270	13	law	MNAM
FNV	Fuero de Navarra	unknown	1270	13	law	MNAM

ID	Title	Author	Date	Century	Genre	Collection
GCI	Grant cronica de Espanya I	Fernández de Heredia	1270	13	history	MNAM
GC3	Grant cronica de Espanya III	Fernández de Heredia	1270	13	history	MNAM
LAT	Livres dou tresor	unknown	1270	13	didactics	MNAM
TFB	Libro de las tahurerías	unknown	1270	13	law	ESMP
TFC	Libro de las tahurerías	unknown	1270	13	law	ESMP
TFD	Libro de las tahurerías	unknown	1270	13	law	ESMP
TFE	Libro de las tahurerías	unknown	1270	13	law	ESMP
TFF	Libro de las tahurerías	unknown	1270	13	law	ESMP
TFG	Libro de las tahurerías	unknown	1270	13	law	ESMP
TFH	Libro de las tahurerías	unknown	1270	13	law	ESMP
TFI	Libro de las tahurerías	unknown	1270	13	law	ESMP
TFJ	Libro de las tahurerías	unknown	1270	13	law	ESMP
TFK	Libro de las tahurerías	unknown	1270	13	law	ESMP
TFL	Libro de las tahurerías	unknown	1270	13	law	ESMP
TFM	Libro de las tahurerías	unknown	1270	13	law	ESMP
TFN	Libro de las tahurerías	unknown	1270	13	law	ESMP
TFO	Libro de las tahurerías	unknown	1270	13	law	ESMP
TFP	Libro de las tahurerías	unknown	1270	13	law	ESMP
TFQ	Libro de las tahurerías	unknown	1270	13	law	ESMP
YMG	Libro de las formas y de las imagenes	Alfonso X el Sabio	1276	13	didactics	PWAX
TFA	Libro de las tahurerías	unknown	1276	13	law	ESMP
AST	Libros del saber de astonomia	Alfonso X el Sabio	1277	13	didactics	PWAX
RAB	Libro del cuadrante señero	Alfonso X el Sabio	1277	13	didactics	PWAX
EE1	Estoria de España	Alfonso X el Sabio	1280	13	history	PWAX
EE2	Estoria de España	Alfonso X el Sabio	1280	13	history	PWAX
GE1	General estoria I	Alfonso X el Sabio	1280	13	history	PWAX
GE4	General estoria IV	Alfonso X el Sabio	1280	13	history	PWAX
CAS	Fueros de Castiella	unknown	1280	13	law	ESMP
CAX	Crónica de Alfonso X	Alfonso X el Sabio	1280	13	history	ESMP
CS4	Crónica de Sancho IV	unknown	1280	13	history	ESMP
REY	Crónica de once reyes	unknown	1280	13	history	ESMP

ID	Title	Author	Date	Century	Genre	Collection
FNA	Fuero general de Navarra	unknown	1280	13	law	FGNA
FNB	Fuero general de Navarra	unknown	1280	13	law	FGNA
FNC	Fuero general de Navarra	unknown	1280	13	law	FGNA
FND	Fuero general de Navarra	unknown	1280	13	law	FGNA
FNE	Fuero general de Navarra	unknown	1280	13	law	FGNA
FNF	Fuero general de Navarra	unknown	1280	13	law	FGNA
FNG	Fuero general de Navarra	unknown	1280	13	law	FGNA
FNH	Fuero general de Navarra	unknown	1280	13	law	FGNA
FNI	Fuero general de Navarra	unknown	1280	13	law	FGNA
FNJ	Fuero general de Navarra	unknown	1280	13	law	FGNA
FNK	Fuero general de Navarra	unknown	1280	13	law	FGNA
FNL	Fuero general de Navarra	unknown	1280	13	law	FGNA
FNM	Fuero general de Navarra	unknown	1280	13	law	FGNA
FNN	Fuero general de Navarra	unknown	1280	13	law	FGNA
FNO	Fuero general de Navarra	unknown	1280	13	law	FGNA
FNP	Fuero general de Navarra	unknown	1280	13	law	FGNA
FNQ	Fuero general de Navarra	unknown	1280	13	law	FGNA
FNR	Fuero general de Navarra	unknown	1280	13	law	FGNA
FNS	Fuero general de Navarra	unknown	1280	13	law	FGNA
ACE	Libros de ajedrez, dados y tablas	Alfonso X el Sabio	1283	13	didactics	PWAX
VLT	Gran conquista de Ultramar	unknown	1290	13	prose	ESMP
CC1	Libro del consejo y del consejero	unknown	1293	13	didactics	ESMP
CC2	Libro del consejo y del consejero	unknown	1293	13	didactics	ESMP
CD1	Castigos y documentos	unknown	1293	13	didactics	ESMP
CD2	Castigos e documentos de Sancho IV	unknown	1293	13	didactics	ESMP
ULT	Gran conquista de Ultramar	unknown	1295	13	prose	ESMP
G2K	General estoria II	Alfonso X el Sabio	1300	14	history	PWAX
GE6	General estoria VI	Alfonso X el Sabio	1300	14	history	PWAX
ACR	Libro de los açores	unknown	1300	14	didactics	ESMP
ARC	Anales de los reyes de Castilla	unknown	1300	14	law	ESMP
CET	Tratado de cetrería	unknown	1300	14	didactics	ESMP

ID	Title	Author	Date	Century	Genre	Collection
DAN	Danza de la muerte	unknown	1300	14	poetry	ESMP
EMZ	De una santa emperatriz que hubo en Roma	unknown	1300	14	prose	ESMP
GLF	Guillelmus falconarius	unknown	1300	14	didactics	ESMP
GRF	Gerardus falconarius	unknown	1300	14	didactics	ESMP
RCY	Recetario	unknown	1300	14	didactics	ESMP
SEM	Proverbios morales de Semtob	unknown	1300	14	didactics	ESMP
SMM	Santa María Madalena	unknown	1300	14	prose	ESMP
LAB	Labores	unknown	1300	14	didactics	CORDE
LA1	Labranzas	unknown	1300	14	didactics	CORDE
CON	Libro del consejo e consejeros	unknown	1300	14	law	CORDE
REC	Receta	unknown	1300	14	medicine	CORDE
PLA	Plantar	unknown	1300	14	didactics	CORDE
PA7	Partida VII Pedro	unknown	1300	14	law	CORDE
FYO	Flor de las ystorias de Orient	Fernández de Heredia	1307	14	history	MNAM
AC1	Leyes del estilo	unknown	1310	14	law	ESMP
AC2	Leyes del estilo	unknown	1310	14	law	ESMP
AC3	Leyes del estilo	unknown	1310	14	law	ESMP
AC5	Fuero real. Leyes del estilo	unknown	1310	14	law	ESMP
BC1	Leyes del estilo	unknown	1310	14	law	ESMP
BC4	Leyes del estilo	unknown	1310	14	law	ESMP
AC6	Leyes de estilo	unknown	1310	14	law	CORDE
BRI	Fuero de Briviesca	unknown	1313	14	law	ESMP
CF1	Libro primero de las Confesiones	Martín Pérez	1315	14	didactics	TMEM
CF2	Segunda parte del Libro de las Confesiones	Martín Pérez	1315	14	didactics	TMEM
CF3	Libro tercero de las Confesiones	Martín Pérez	1315	14	didactics	TMEM
LEO	Sumas de la historia troyana	Leomarte	1320	14	history	ESMP
TRE	Libro de las tres creencias	Alfonso de Valladolid	1320	14	religion	ESMP
MCR	Crónica abreviada	Juan Manuel	1321	14	history	ESMP
HCP	Historia del caballero Plácidas	unknown	1325	14	prose	ESMP
MCE	Libro del caballero y del escudero	Juan Manuel	1326	14	didactics	ESMP
MCA	Libro de la caza	Juan Manuel	1330	14	didactics	ESMP

ID	Title	Author	Date	Century	Genre	Collection
MES	Libro de los estados	Juan Manuel	1330	14	didactics	ESMP
FN1	Fuero general de Navarra	unknown	1330	14	law	CORDE
LUC	Libro del conde Lucanor	Juan Manuel	1331	14	prose	ESMP
MIN	Libro infinido	Juan Manuel	1336	14	didactics	ESMP
LBG	Libro de buen amor-G	Juan Ruiz	1340	14	poetry	ESMP
LBS	Libro de buen amor-S	Juan Ruiz	1340	14	poetry	ESMP
LBT	Libro de buen amor-T	Juan Ruiz	1340	14	poetry	ESMP
LC1	Libro del conocimiento de todos los reinos	unknown	1340	14	history	ESMP
LC2	Libro del conocimiento de todos los reinos	unknown	1340	14	history	ESMP
LC3	Libro del conocimiento de todos los reinos	unknown	1340	14	history	ESMP
MAR	Libro de las armas	Juan Manuel	1340	14	didactics	ESMP
MON	Libro de la montería	Alfonso XI	1340	14	didactics	ESMP
MAS	Tratado de la asunción	Juan Manuel	1342	14	religion	ESMP
CRN	Crónica de 1344	unknown	1344	14	history	ESMP
AO2	Ordenamiento de Alcalá	unknown	1348	14	law	ESMP
FVA	Fuero Viejo de Alcalá	Alfonso XI	1348	14	law	ESMP
MRP	Libro de Marco Polo	unknown	1350	14	prose	MNAM
NO1	Fuero de la Novenera	unknown	1350	14	law	MNAM
MEN	Tratado de menescalcia	Aluares de Salamiellas	1350	14	didactics	ESMP
RAH	Biblia romanceada	unknown	1350	14	religion	ESMP
VSC	Vida de Santa Catalina	unknown	1350	14	prose	ESMP
VIS	Visita y consejo de médicos	Estephano de Sevilla	1350	14	medicine	HMSM
MDR	Mocedades de Rodrigo	unknown	1360	14	prose	ESMP
CQ1	Cronica de los conqueridores I	Fernández de Heredia	1370	14	history	MNAM
MOR	Cronica de Morea	Fernández de Heredia	1370	14	history	MNAM
ORS	Historia contra los paganos de Orosio	unknown	1370	14	history	MNAM
PL1	Plutarch I	unknown	1370	14	history	MNAM
PL2	Plutarch II	unknown	1370	14	history	MNAM
PL3	Plutarch III	unknown	1370	14	history	MNAM
RAM	Rams de flors o Libro de actoridades	unknown	1370	14	prose	MNAM
SEC	Secreto secretorum	unknown	1370	14	didactics	MNAM

ID	Title	Author	Date	Century	Genre	Collection
TRY	Historia troyana	unknown	1370	14	history	MNAM
TUC	Tucidides	unknown	1370	14	history	MNAM
TUY	Cronica del tudense	unknown	1370	14	history	MNAM
OA1	Ordenamiento de Alcalá	Alfonso XI	1370	14	law	ESMP
TRS	Cuento de Tristán de Leonis	unknown	1370	14	prose	ESMP
HJA	Historia de Jerusalén abreviada de Jacobo de Vitriaco	unknown	1370	14	prose	TMEM
EUT	Breviarium ab urbe condita	Fernández de Heredia	1377	14	history	MNAM
EMP	Cronica de los emperadores	Fernández de Heredia	1380	14	history	MNAM
MAN	Viajes de Juan de Mandevilla	unknown	1380	14	prose	MNAM
BUR	Fuero de Burgos	unknown	1380	14	law	ESMP
PAL	Libro de Palladio	unknown	1380	14	didactics	ESMP
CQ2	Cronica de los conqueridores II	Fernández de Heredia	1385	14	history	MNAM
ICR	Crónica de San Isidoro	unknown	1385	14	history	ESMP
IHG	Historia de los godos de San Isidoro	unknown	1385	14	history	ESMP
LCA	Libro de la caça de las aues	Pero López de Ayala	1386	14	didactics	ESMP
EUG	Cronica general de Espana	Garcia de Eugui	1390	14	history	MNAM
G5R	General estoria V	Alfonso X el Sabio	1400	15	history	PWAX
AMZ	Amazonas	unknown	1400	15	history	ESMP
APO	Libro de Apolonio	unknown	1400	15	prose	ESMP
ATA	Atalaya de las corónicas	Arcipreste de Talavera	1400	15	history	ESMP
AUG	Meditationes de Pseudo-Augustine, De infantia Salvatoris	unknown	1400	15	didactics	ESMP
BAE	Cancionero de Baena	unknown	1400	15	poetry	ESMP
BRA	De la reformación de la ánima	unknown	1400	15	didactics	ESMP
BRN	Biblia Romanceada	unknown	1400	15	religion	ESMP
BSM	Bendita Santa Marta	unknown	1400	15	religion	ESMP
CCC	Cancionero castellano y catalán de París	unknown	1400	15	poetry	ESMP
CCM	Cuento del emperador Carlos Maynes	unknown	1400	15	prose	ESMP
CDS	Castigos y dotrinas	unknown	1400	15	didactics	ESMP
CGS	Castigos y documentos	unknown	1400	15	didactics	ESMP
CMN	Arte complida de cirugía	unknown	1400	15	medicine	ESMP
CUZ	Libro del kuzari	unknown	1400	15	religion	ESMP

ID	Title	Author	Date	Century	Genre	Collection
EOR	El emperador Otas de Roma	unknown	1400	15	prose	ESMP
ERG	Estoria del rey Gujllelme	unknown	1400	15	prose	ESMP
GYS	Generaciones y semblanzas	Pérez de Guzmán	1400	15	prose	ESMP
NO2	Fueros de la Novenera	unknown	1400	15	law	ESMP
P13	Cancionero de Salvá	unknown	1400	15	poetry	ESMP
PN2	Cancionero castellano de París	unknown	1400	15	poetry	ESMP
PN5	Cancionero castellano de París	unknown	1400	15	poetry	ESMP
PN9	Cancionero castellano de París	unknown	1400	15	poetry	ESMP
PRS	Cancionero de París	unknown	1400	15	poetry	ESMP
RHE	De la rhetorica	unknown	1400	15	didactics	ESMP
DTM	Diccionario Eclesiástico o Teológico-Moral	unknown	1400	15	religion	TMEM
LDA	Libro de Astrologia	unknown	1400	15	didactics	TMEM
PSL	Proverbios de Seneca llamados vicios y virtudes	unknown	1400	15	didactics	TMEM
ALV	Regimiento contra la peste	Fernando Alvarez	1400	15	medicine	HMSM
TAM	Historia del gran Tamerlín	Ruy González de Clavijo	1410	15	prose	ESMP
SL2	Suma de las corónicas	Pablo de Santa María	1412	15	history	ESMP
PER	Mostrador e enseñador de los turbados	Moises ben Maimon	1420	15	didactics	ESMP
OFI	De officiis de Cicerón	unknown	1422	15	law	ESMP
CIS	Arte cisoria	Enrique de Villena	1423	15	didactics	ESMP
LPA	Libro de las paradojas	Alfonso de Madrigal	1437	15	didactics	TMEM
CBO	Corbacho	Arcipreste de Talavera	1438	15	didactics	ESMP
ARB	Árbol de Batallas	Antón Zorita	1441	15	history	TMEM
DFN	Defensa de virtuossas mugeres	Diego de Valera	1445	15	didactics	ESMP
DIV	Tratado de adivinar y de magia	Lope de Barrientos	1445	15	didactics	ESMP
DV2	Tratado de divinanza	Lope de Barrientos	1445	15	didactics	ESMP
DON	Libro de las donas	Francesc Eiximenis	1448	15	didactics	ESMP
AXI	Axioco	Pedro Díaz de Toledo	1450	15	prose	ESMP
EVG	Libro de cetrería de Evangelista	unknown	1450	15	didactics	ESMP
JAM	Texto jurídico aljamiado	unknown	1450	15	law	ESMP
OLI	Enrique fi de Oliva	unknown	1450	15	prose	ESMP
SER	Sermones contra los iudios e moros	unknown	1450	15	didactics	ESMP

ID	Title	Author	Date	Century	Genre	Collection
LAA	Libro de amor y amiçiçia de latin	Alfonso de Madrigal	1450	15	prose	TMEM
LVC	Libro de Vejecio de la caballeria	Alfonso de San Cristóbal	1450	15	didactics	TMEM
TVV	Tratado de viçios y virtudes	unknown	1450	15	didactics	TMEM
OVM	Morales de Ovidio	Pierre Bersuire	1452	15	didactics	ESMP
INV	Invencionario	Alfonso de Toledo	1453	15	law	ESMP
JPM	Carta de Juan II a la condesa Juana Pimentel	Juan II	1453	15	letters	ESMP
LUN	Carta de Juan II a D. Alvaro de Luna	Juan II	1453	15	letters	ESMP
EIV	Carta de Juan II a Enrique IV	Juan II	1454	15	letters	ESMP
MUS	Tratado de la música	unknown	1460	15	didactics	ESMP
P10	Cancionero de París	unknown	1465	15	poetry	ESMP
TA1	Tratado de las armas	Diego de Valera	1465	15	didactics	ESMP
TA2	Tratado de las armas	Diego de Valera	1465	15	didactics	ESMP
TAL	Tratado del Alborayque	unknown	1465	15	didactics	ESMP
MPM	Mapamundi atribuido a San Isidoro	unknown	1467	15	didactics	TMEM
JCE	Juego de Claudio emperador	unknown	1470	15	didactics	ESMP
TR1	Triunfo de amor	Juan de Flores	1475	15	prose	ESMP
TR2	Triunfo de amor	Juan de Flores	1475	15	prose	ESMP
VES	De como se ha de ocupar vna sennora cada dia	Hernando de Talavera	1477	15	didactics	TMEM
G&G	Grimalte y Gradissa	Juan de Flores	1480	15	prose	ESMP
RRA	Razonamiento de las Reales Armas de los Católicos Reyes	Antonio de Villalpando	1480	15	history	TMEM
GOW	Confession Amantis de John Gower	unknown	1483	15	didactics	ESMP
DBM	Arte y doctrina de bien morir y breve confesionario	unknown	1484	15	religion	ESMP
MTV	Ordenanzas reales	Díaz de Montalvo	1484	15	law	ESMP
LTR	Letras	Hernando del Pulgar	1485	15	didactics	ESMP
MGO	Coplas y glosas de Mingo Revulgo	unknown	1485	15	poetry	ESMP
CLV	Claros varones de Castilla	unknown	1486	15	history	ESMP
APL	Libro de Apolonio	unknown	1488	15	prose	ESMP
CAR	Cárcel de amor	unknown	1488	15	prose	ESMP
MEL	Historia de la linda Melosina	Jean D'Arras	1489	15	prose	ESMP
AYL	Arnalte y Lucenda	unknown	1490	15	prose	ESMP
CTY	Crónica Troyana	unknown	1490	15	history	ESMP

ID	Title	Author	Date	Century	Genre	Collection
IMI	Imitatio Christi	Pseudo-Jean Gerson	1490	15	religion	ESMP
VAS	La estoria del noble Vaspasiano	unknown	1490	15	prose	ESMP
Y88	Esopete ystoriado	unknown	1490	15	prose	ESMP
Y89	Ysopete ystoriado	unknown	1490	15	prose	ESMP
SPO	Siete Partidas	Alfonso X el Sabio	1491	15	law	ESMP
DHL	Dictionarium hispano-latinum	Antonio de Nebrija	1492	15	didactics	ESMP
GRM	Gramática castellana	Antonio de Nebrija	1492	15	didactics	ESMP
ABC	Exemplario por ABC	unknown	1493	15	prose	ESMP
CAP	Exemplario contra los engaños y peligros del mundo	Juan de Capua	1493	15	prose	ESMP
COL	Carta de Colón	unknown	1493	15	letters	ESMP
SL3	Conjuración de Catalina	Vidal de Noya	1493	15	history	ESMP
SLC	Conjuración de Catalina	Vidal de Noya	1493	15	history	ESMP
SLI	Guerra de Jugurtha	Vidal de Noya	1493	15	history	ESMP
SLM	Guerra de Jugurtha	unknown	1493	15	history	ESMP
CAU	Tratado de cirugía	Guido de Cauliaco	1493	15	medicine	HMSM
LIM	De las ilustres mujeres de Boccaccio	unknown	1494	15	didactics	ESMP
CDP	Caída de prinçipes de Boccacio	unknown	1495	15	prose	ESMP
DLH	Dictionarium latino-hispanicum	Antonio de Nebrija	1495	15	didactics	ESMP
ENC	Cancionero de las obras de Juan del Encina	Juan del Encina	1496	15	poetry	ESMP
BLA	Libro de los blasones de reyes y grandes señores	Hernández de Mendoza	1496	15	prose	TMEM
GLX	Glosa sobre lux bella	Marcos Durán	1498	15	didactics	ESMP
LUX	Lux bella	Marcos Durán	1498	15	didactics	ESMP
C01	La Celestina	unknown	1499	15	prose	ECTC
COR	Coronación de Juan de Mena	unknown	1499	15	poetry	ESMP
HER	Los doze trabajos de Hércules	Enrique de Villena	1499	15	didactics	ESMP
MAE	Muestra de las antigüedades de España	Antonio de Nebrija	1499	15	history	ESMP
OLV	Oliveros de Castilla	unknown	1499	15	prose	ESMP
C02	La Celestina	unknown	1500	16	prose	ECTC
C07	La Celestina	unknown	1500	16	prose	ECTC
C08	La Celestina	unknown	1500	16	prose	ECTC
C17	La Celestina	unknown	1500	16	prose	ECTC

ID	Title	Author	Date	Century	Genre	Collection
C21	La Celestina	unknown	1500	16	prose	ECTC
GOD	Anales de los reyes godos	Jiménez de Rada	1500	16	law	ESMP
HLC	Abreviación del halconero	unknown	1500	16	history	ESMP
RHJ	Romance del rey que no hace justicia	unknown	1500	16	poetry	ESMP
STR	Strategematon	Frontino	1500	16	didactics	ESMP
CHA	Tratado nuevo	Alvarez Chanca	1500	16	medicine	HMSM
CHS	Compendio de la humana salud	Johannes de Ketham	1500	16	medicine	HMSM
CIR	Cirugía rimada	Diego de Covo	1500	16	medicine	HMSM
CMY	Cirugía mayor	Lanfranco de Milán	1500	16	medicine	HMSM
ESP	Espejo de medicina	Alfonso Chirino	1500	16	medicine	HMSM
FIE	Tratado de las fiebres	Isaac Israelí	1500	16	medicine	HMSM
FLO	Suma de la flor de cirugía	Fernando de Córdoba	1500	16	medicine	HMSM
FOR	Tratado útil	Licenciado Fores	1500	16	medicine	HMSM
GEN	Tratado de la generación de la criatura Madrid	unknown	1500	16	medicine	HMSM
GIL	El libro de recetas	Gilberto	1500	16	medicine	HMSM
GOM	Compendio de medicina	Gómez de Salamanca	1500	16	medicine	HMSM
GOR	Lilio de medicina	Bernardo de Gordonio	1500	16	medicine	HMSM
GRA	Cirugía	unknown	1500	16	medicine	HMSM
LOL	Libro de los olios	unknown	1500	16	medicine	HMSM
MAC	Macer herbolario	unknown	1500	16	medicine	HMSM
MLS	De las melecinas	unknown	1500	16	medicine	HMSM
RES	Recetas	unknown	1500	16	medicine	HMSM
ROM	Propiedades del romero	Gómez de Salamanca	1500	16	medicine	HMSM
SDM	Secretos de la medicina	Juan Enríquez	1500	16	medicine	HMSM
SEV	Sevillana medicina	Juan de Aviñón	1500	16	medicine	HMSM
SUM	Sumario de la medicina	López de Villalobos	1500	16	medicine	HMSM
TED	Cirugía	Tedrico	1500	16	medicine	HMSM
TES	Tesoro de la medicina	unknown	1500	16	medicine	HMSM
TPH	Tratado de Phisonomía	unknown	1500	16	medicine	HMSM
TRM	Tratado de la patología general	unknown	1500	16	medicine	HMSM
TRP	Tratado de la epidemia y pestilencia	Velasco de Taranto	1500	16	medicine	HMSM

ID	Title	Author	Date	Century	Genre	Collection
C03	La Celestina	unknown	1501	16	prose	ECTC
TRI	Libro de don Tristán de Leonís	unknown	1501	16	prose	ESMP
GAH	Obra de agricultura	Alonso de Herrera	1504	16	didactics	ESMP
LTO	Leyes de Toro	unknown	1505	16	law	ESMP
C10	La Celestina	unknown	1507	16	prose	ECTC
C04	La Celestina	unknown	1511	16	prose	ECTC
C09	La Celestina	unknown	1512	16	prose	ECTC
CPC	Crónica particular del Cid	unknown	1512	16	history	ESMP
C05	La Celestina	unknown	1513	16	prose	ECTC
CHI	Menor daño de la medicina	Alfonso Chirino	1513	16	medicine	HMSM
C11	La Celestina	unknown	1514	16	prose	ECTC
BOT	Compendio de los boticarios	Saladino da Ascoli	1515	16	medicine	HMSM
DDH	Tabla de la diversidad de los días y horas	unknown	1517	16	didactics	ESMP
LZ1	Lazarillo de Tormes	unknown	1517	16	prose	ESMP
LZ2	Lazarillo de Tormes	unknown	1517	16	prose	ESMP
LZ3	Lazarillo de Tormes	unknown	1517	16	prose	ESMP
ORT	Reglas de orthographia	Antonio de Nebrija	1517	16	didactics	ESMP
C06	La Celestina	unknown	1518	16	prose	ECTC
C12	La Celestina	unknown	1518	16	prose	ECTC
C13	La Celestina	unknown	1518	16	prose	ECTC
M18	Libro de medecina llamado macer	Maestre Gil	1518	16	medicine	HMSM
MAG	Pierres e Magalona	Bernardo de Treviez	1519	16	prose	ESMP
M19	Libro de medecina llamado macer	Maestre Gil	1519	16	medicine	HMSM
P&V	La ystoria del noble cauallero Paris	unknown	1524	16	prose	ESMP
C14	La Celestina	unknown	1525	16	prose	ECTC
C15	La Celestina	unknown	1525	16	prose	ECTC
CEGH	Diálogos	Vives	1525	16	didactics	Cervantes
C16	La Celestina	unknown	1526	16	prose	ECTC
CDC	Crónica del Cid	unknown	1526	16	prose	ESMP
C18	La Celestina	unknown	1528	16	prose	ECTC
C19	La Celestina	unknown	1529	16	prose	ECTC

ID	Title	Author	Date	Century	Genre	Collection
C20	La Celestina	unknown	1540	16	prose	ECTC
CBN	Libro del arte de las comadres o madrinas	Damián Carbón	1541	16	medicine	HMSM
YSL	Tractado llamado fructo de todos los auctos	Ruy Díaz de Ysla	1542	16	medicine	HMSM
GHOZ	Naufragios	Núñez Cabeza de Vaca	1542	16	prose	Gutenberg
LAH	Libro de la Anathomia del hombre	Montaña de Monserrate	1551	16	medicine	HMSM
GHIU	La vida de Lazarillo de tormes y de sus fortunas	unknown	1554	16	prose	Gutenberg
GHLO	Segunda parte de la crónica del Perú	Cieza de León	1554	16	history	Gutenberg
LAG	Discvrso breve, sobre la cvra y preservacion de la pestilencia	unknown	1556	16	didactics	ESMP
SDC	Secretos de chirurgia	Arias de Benavides	1567	16	medicine	HMSM
TBC	Tratado breve y copipendioso	Martínez de Castrillo	1570	16	medicine	HMSM
IND	Discvrsos de las cosas Aromaticas	Juan Fragoso	1572	16	didactics	ESMP
MRD	Primera y Segvnda y Tercera Partes de la Historia Medicinal	Nicolás Monardes	1580	16	medicine	ESMP
PAR	Libro intitulado del parto humano	Francisco Núñez	1580	16	medicine	HMSM
SAN	Tratado de la utilidad de la sangría en las viruelas	Fernando de Valdés	1583	16	medicine	HMSM
TDM	Tractado del uso de las mugeres	Francisco Núñez	1586	16	medicine	HMSM
BGM	Enchiridion o manual contra la gota	Gímez Miedes	1589	16	medicine	HMSM
DJC	Discurso y despertador preservativo de enfermedades	Juan Cornejo	1594	16	medicine	HMSM
DPP	Discurso particular preservativo de la Gota	Juan Cornejo	1594	16	medicine	HMSM
BTP	Breve tratado de peste	Antonio Pérez	1598	16	medicine	HMSM
EXP	Libro de experimentos medicos, faciles y verdaderos	Jerónimo Soriano	1598	16	medicine	HMSM
MFF	Tratado de la peste	Marsilio Ficino	1598	16	medicine	HMSM
ALG	Para el aprovechamiento y examen de los algebristas	Luis Mercado	1599	16	medicine	HMSM
ZAM	Orden para la cvra y preservacion de las secas y carbuncos	Zamudio de Alfaro	1599	16	medicine	HMSM
LEB	Libro que trata de lo enfermedad de las bubas	Pedro de Torres	1600	17	medicine	HMSM
GHKZ	Don Quijote	Cervantes	1604	17	prose	Gutenberg
GHLN	Viage al Parnaso, La Numancia y El Trato de Argel	Cervantes	1614	17	poetry	Gutenberg
GHLM	Novelas y teatro	Cervantes	1616	17	prose	Gutenberg
CEFT	Fuente Ovejuna	Lope	1618	17	drama	Cervantes
GHOT	El remedio en la desdicha y El mejor alcalde, el rey	Lope	1620	17	drama	Gutenberg
GHOV	Expedición de Catalanes y Argoneses al Oriente	Moncada	1623	17	history	Gutenberg
GHST	Historia de la vida del Buscón, llamado Don Pablos	Quevedo	1626	17	prose	Gutenberg

ID	Title	Author	Date	Century	Genre	Collection
GHOU	La moza de cántaro	Lope	1627	17	drama	Gutenberg
GHIM	Los favores del mundo	Alarcón	1628	17	drama	Gutenberg
CEFS	La vida es sueño	Calderón	1636	17	drama	Cervantes
CEFR	El alcalde de Zalamea	Calderón	1651	17	drama	Cervantes
GHNV	La fe triunfante en quatro autos	Garau	1691	17	didactics	Gutenberg
GHQZ	Relacion historial de las misiones de indios chiquitos	Patricio Fernández	1726	18	history	Gutenberg
GIJS	Logica	Piquer	1747	18	didactics	Gutenberg
CEFQ	Cartas marruecas	Cadalso	1773	18	letters	Cervantes
GHOR	Fábulas literarias	Iriarte	1782	18	poetry	Gutenberg
CEFW	Obras dramáticas	Quintana	1800	19	drama	Cervantes
CEFU	El sí de las niñas	Fernández de Moratín	1806	19	drama	Cervantes
GHIW	Filosofia Fundamental I-IV	Balmes	1818	19	didactics	Gutenberg
GHNZ	Contigo Pan y Cebolla	Gorostiza	1833	19	prose	Gutenberg
CEFZ	Don Álvaro o la fuerza del sino	Rivas	1835	19	drama	Cervantes
GHNW	El trovador	García Gutiérrez	1836	19	drama	Gutenberg
GHOS	Artículos selectos	Larra	1836	19	newspaper	Gutenberg
GHLU	El Estudiante de Salamanca	Espronceda	1837	19	poetry	Gutenberg
GHOP	Los Amantes de Teruel	Hartzenbusch	1837	19	drama	Gutenberg
GHIS	De Las Islas Filipinas	Alvarez y Tejero	1842	19	history	Gutenberg
CEGI	Don Juan Tenorio	Zorrilla	1844	19	drama	Cervantes
GHOW	Descripcion é historia del castillo de la aljafería	Nougués Secall	1846	19	history	Gutenberg
GHIT	Historia de la célebre Reina de España Doña Juana	unknown	1848	19	history	Gutenberg
GHIZ	Filosofia Fundamental, tomo III	Balmes	1848	19	didactics	Gutenberg
GHJK	Filosofia fundamental, tomo I	Balmes	1848	19	didactics	Gutenberg
GHJL	Filosofia fundamental, tomo IV	Balmes	1848	19	didactics	Gutenberg
GHIV	El Criterio	Balmes	1849	19	didactics	Gutenberg
GHKU	La gaviota	Caballero	1849	19	prose	Gutenberg
GHIN	Novelas cortas	Alarcón	1854	19	prose	Gutenberg
GHMO	El cocinero de su majestad	Fernández González	1858	19	prose	Gutenberg
GHIO	Viajes por España	Alarcón	1860	19	prose	Gutenberg
GHLS	Fortuna	Pérez Escrich	1860	19	prose	Gutenberg

ID	Title	Author	Date	Century	Genre	Collection
GHSW	Más vale maña que fuerza	Tamayo y Baus	1860	19	drama	Gutenberg
GHJM	Un paseo por Paris, retratos al natural	Barcia	1863	19	prose	Gutenberg
GHRT	Escenas Montañesas	Pereda	1864	19	prose	Gutenberg
GHJU	Historia de Teruel	Blasco y Val	1870	19	history	Gutenberg
GHMZ	La Fontana de Oro	Pérez Galdos	1870	19	prose	Gutenberg
CEFP	Rimas y leyendas	Bécquer	1871	19	poetry	Cervantes
GHRV	Los Hombres de Pro	Pereda	1872	19	prose	Gutenberg
GHNT	Trafalgar	Pérez Galdos	1873	19	prose	Gutenberg
GHIL	El sombrero de tres picos	Alarcón	1874	19	prose	Gutenberg
GHMP	El manco de Lepanto	Fernández González	1874	19	prose	Gutenberg
GHMR	Cádiz	Pérez Galdos	1874	19	prose	Gutenberg
GIJK	Pepita Jiménez	Valera	1874	19	prose	Gutenberg
GHKW	Impresiones, Poesías	Campo-Arana	1876	19	poetry	Gutenberg
GHNS	Torquemada en la hoguera	Pérez Galdos	1876	19	prose	Gutenberg
GHUV	El Comendador Mendoza	Valera	1876	19	prose	Gutenberg
GHWZ	Pasarse de listo	Valera	1876	19	prose	Gutenberg
GHMQ	Bailén	Pérez Galdos	1878	19	prose	Gutenberg
GHNO	Marianela	Pérez Galdos	1878	19	prose	Gutenberg
GHJQ	Suma y narracion de los Incas	Betánzos	1880	19	prose	Gutenberg
GHIK	El Capitán Veneno	Alarcón	1881	19	prose	Gutenberg
GHMW	La desheredada	Pérez Galdos	1881	19	prose	Gutenberg
GHQW	Un viaje de novios	Pardo Bazán	1881	19	prose	Gutenberg
GHTW	Cuentos y diálogos	Valera	1882	19	prose	Gutenberg
GIJQ	Lázaro casi novela	Picón	1882	19	prose	Gutenberg
GHPW	Marta y María	Palacio Valdés	1883	19	prose	Gutenberg
GHTV	Algo de todo	Valera	1883	19	prose	Gutenberg
GHLP	La Regenta	Clarín	1884	19	prose	Gutenberg
GHLZ	El infierno del amor leyenda fantástica	Fernández González	1884	19	poetry	Gutenberg
GHMV	La de Bringas	Pérez Galdos	1884	19	prose	Gutenberg
GHNR	Tormento	Pérez Galdos	1884	19	prose	Gutenberg
GHPQ	Aguas fuertes	Palacio Valdés	1884	19	prose	Gutenberg

ID	Title	Author	Date	Century	Genre	Collection
CEFV	Cuentos de amor	Pardo Bazán	1885	19	prose	Cervantes
GHNU	Un faccioso más y algunos frailes menos	Pérez Galdos	1885	19	prose	Gutenberg
GHTU	José	Palacio Valdés	1885	19	prose	Gutenberg
GHPZ	Riverita	Palacio Valdés	1886	19	prose	Gutenberg
GHQV	Los pazos de Ulloa	Pardo Bazán	1886	19	prose	Gutenberg
GHIP	Viajes por Filipinas: De Manila á Tayabas	Álvarez Guerra	1887	19	prose	Gutenberg
GHIQ	Viajes por Filipinas: De Manila á Marianas	Álvarez Guerra	1887	19	prose	Gutenberg
GHMU	Fortunata y Jacinta dos historias de casadas	Pérez Galdos	1887	19	prose	Gutenberg
GIJO	El enemigo	Picón	1887	19	prose	Gutenberg
GHMN	Amparo o Memorias de un loco	Fernández González	1888	19	prose	Gutenberg
GHQS	El cuarto poder	Palacio Valdés	1888	19	prose	Gutenberg
GHRU	La Montálvez	Pereda	1888	19	prose	Gutenberg
GHSU	Tres Comedias Modernas en un acto y en prosa	Barranco	1888	19	drama	Gutenberg
CEFO	Novelas cortas	Asensi	1889	19	prose	Cervantes
GHPU	La hermana San Sulpicio	Palacio Valdés	1889	19	prose	Gutenberg
GHLQ	Su único hijo	Clarín	1890	19	prose	Gutenberg
GHLW	Estudios históricos del reinado de Felipe II	Fernández Duro	1890	19	history	Gutenberg
GHQT	La Espuma	Palacio Valdés	1890	19	prose	Gutenberg
GHRS	Al primer vuelo	Pereda	1890	19	prose	Gutenberg
GHUW	El Superhombre y otras novedades	Valera	1890	19	prose	Gutenberg
GIJN	Dulce y sabrosa	Picón	1891	19	prose	Gutenberg
GHQU	La Fe	Palacio Valdés	1892	19	prose	Gutenberg
GHPS	El maestrante	Palacio Valdés	1893	19	prose	Gutenberg
GHPT	El origen del pensamiento	Palacio Valdés	1893	19	prose	Gutenberg
GHIJ	Historia y Geografía de Mindanao	Nieto Aguilar	1894	19	history	Gutenberg
GHJV	Arroz y tartana	Blasco Ibáñez	1894	19	prose	Gutenberg
GHPR	El idilio de un enfermo	Palacio Valdés	1894	19	prose	Gutenberg
GHRW	Peñas arriba	Pereda	1894	19	prose	Gutenberg
GHVW	Juanita La Larga	Valera	1895	19	prose	Gutenberg
GIJM	Cuentos de mi tiempo	Picón	1895	19	prose	Gutenberg
GIJR	Tres mujeres	Picón	1896	19	prose	Gutenberg

ID	Title	Author	Date	Century	Genre	Collection
GHMS	Doña Perfecta	Pérez Galdos	1897	19	prose	Gutenberg
GHNQ	Misericordia	Pérez Galdos	1897	19	prose	Gutenberg
GHTZ	Doña Luz	Valera	1897	19	prose	Gutenberg
GHUZ	Genio y figura	Valera	1897	19	prose	Gutenberg
GHKM	La Barraca	Blasco Ibáñez	1898	19	prose	Gutenberg
GHLR	Pequeñeces	Coloma	1898	19	prose	Gutenberg
GHVZ	Morsamor	Valera	1899	19	prose	Gutenberg
GHMT	Electra	Pérez Galdos	1901	20	drama	Gutenberg
GIJP	La vistosa	Picón	1901	20	prose	Gutenberg
GHKN	La Catedral	Blasco Ibáñez	1903	20	prose	Gutenberg
GHNP	Mariucha	Pérez Galdos	1903	20	prose	Gutenberg
GHSZ	Antonio Azorín	Trinidad Martínez Ruiz	1903	20	prose	Gutenberg
GHJW	Entre naranjos	Blasco Ibáñez	1904	20	prose	Gutenberg
GHKO	El intruso	Blasco Ibáñez	1904	20	prose	Gutenberg
GHPV	Los Puritanos, y otros cuentos	Palacio Valdés	1904	20	prose	Gutenberg
GHKR	La bodega	Blasco Ibáñez	1905	20	prose	Gutenberg
GHQR	Tristán o el pesimismo	Palacio Valdés	1906	20	prose	Gutenberg
GIJL	Romance de lobos, comedia barbara	Valle-Inclán	1907	20	drama	Gutenberg
GHJT	Los muertos mandan	Blasco Ibáñez	1908	20	prose	Gutenberg
GHOQ	El pecado y la noche	Hoyos y Vinent	1912	20	prose	Gutenberg
GHJR	Los argonautas	Blasco Ibáñez	1913	20	prose	Gutenberg
GHIR	Doña Clarines y Mañana de Sol	Quintero	1915	20	drama	Gutenberg
GHJS	Los cuatro jinetes del apocalipsis	Blasco Ibáñez	1916	20	prose	Gutenberg
GHKS	La condenada	Blasco Ibáñez	1916	20	prose	Gutenberg
GHJP	Tres Comedias	Benavente	1918	20	prose	Gutenberg
GHJO	Zalacaín el Aventurero	Baroja	1919	20	prose	Gutenberg
GHJZ	Mare nostrum	Blasco Ibáñez	1919	20	prose	Gutenberg
GHKL	Sangre y arena	Blasco Ibáñez	1919	20	prose	Gutenberg
GHLV	Novelas y cuentos	Calderón	1919	20	prose	Gutenberg
GHJN	Las inquietudes de Shanti Andia	Baroja	1920	20	prose	Gutenberg
GHKV	La rana viajera	Camba	1920	20	prose	Gutenberg

ID	Title	Author	Date	Century	Genre	Collection
GHRZ	Belarmino y Apolonio	Pérez de Ayala	1920	20	prose	Gutenberg
GHKQ	El préstamo de la difunta	Blasco Ibáñez	1921	20	prose	Gutenberg
GHKP	El paraiso de las mujeres	Blasco Ibáñez	1922	20	prose	Gutenberg
GHKT	La Tierra de Todos	Blasco Ibáñez	1922	20	prose	Gutenberg
GHLT	La Niña de Luzmela	Espina	1922	20	prose	Gutenberg
GHSV	La Puerta de Bronce y Otros Cuentos	Romero de Terreros	1922	20	prose	Gutenberg
a29	unknown	unknown	1978-1995	20	prose	Lexesp
d4	unknown	unknown	1978-1995	20	prose	Lexesp
e15	unknown	unknown	1978-1995	20	prose	Lexesp
n2.2	unknown	unknown	1978-1995	20	prose	Lexesp
t10	unknown	unknown	1978-1995	20	prose	Lexesp
t30	unknown	unknown	1978-1995	20	prose	Lexesp
t48	unknown	unknown	1978-1995	20	prose	Lexesp
a1	unknown	unknown	1978-1995	20	prose	Lexesp
a3	unknown	unknown	1978-1995	20	prose	Lexesp
d5.1	unknown	unknown	1978-1995	20	prose	Lexesp
e16	unknown	unknown	1978-1995	20	prose	Lexesp
n2.3	unknown	unknown	1978-1995	20	prose	Lexesp
t11	unknown	unknown	1978-1995	20	prose	Lexesp
t31	unknown	unknown	1978-1995	20	prose	Lexesp
t49	unknown	unknown	1978-1995	20	prose	Lexesp
a10	unknown	unknown	1978-1995	20	prose	Lexesp
a30	unknown	unknown	1978-1995	20	prose	Lexesp
d5.2	unknown	unknown	1978-1995	20	prose	Lexesp
e17	unknown	unknown	1978-1995	20	prose	Lexesp
n3	unknown	unknown	1978-1995	20	prose	Lexesp
t12	unknown	unknown	1978-1995	20	prose	Lexesp
t32	unknown	unknown	1978-1995	20	prose	Lexesp
t5	unknown	unknown	1978-1995	20	prose	Lexesp
a11	unknown	unknown	1978-1995	20	prose	Lexesp
a4	unknown	unknown	1978-1995	20	prose	Lexesp

ID	Title	Author	Date	Century	Genre	Collection
d5.3	unknown	unknown	1978-1995	20	prose	Lexesp
e18.1	unknown	unknown	1978-1995	20	prose	Lexesp
n4	unknown	unknown	1978-1995	20	prose	Lexesp
t13	unknown	unknown	1978-1995	20	prose	Lexesp
t33	unknown	unknown	1978-1995	20	prose	Lexesp
t50	unknown	unknown	1978-1995	20	prose	Lexesp
a12	unknown	unknown	1978-1995	20	prose	Lexesp
a5	unknown	unknown	1978-1995	20	prose	Lexesp
dc1	unknown	unknown	1978-1995	20	prose	Lexesp
e18.2	unknown	unknown	1978-1995	20	prose	Lexesp
n5	unknown	unknown	1978-1995	20	prose	Lexesp
t14	unknown	unknown	1978-1995	20	prose	Lexesp
t34.1	unknown	unknown	1978-1995	20	prose	Lexesp
t51	unknown	unknown	1978-1995	20	prose	Lexesp
a13	unknown	unknown	1978-1995	20	prose	Lexesp
a6	unknown	unknown	1978-1995	20	prose	Lexesp
dc10	unknown	unknown	1978-1995	20	prose	Lexesp
e18.3	unknown	unknown	1978-1995	20	prose	Lexesp
n6	unknown	unknown	1978-1995	20	prose	Lexesp
t15	unknown	unknown	1978-1995	20	prose	Lexesp
t34.2	unknown	unknown	1978-1995	20	prose	Lexesp
t52	unknown	unknown	1978-1995	20	prose	Lexesp
a14	unknown	unknown	1978-1995	20	prose	Lexesp
a7	unknown	unknown	1978-1995	20	prose	Lexesp
dc2	unknown	unknown	1978-1995	20	prose	Lexesp
e2	unknown	unknown	1978-1995	20	prose	Lexesp
n7	unknown	unknown	1978-1995	20	prose	Lexesp
t16	unknown	unknown	1978-1995	20	prose	Lexesp
t34.3	unknown	unknown	1978-1995	20	prose	Lexesp
t53	unknown	unknown	1978-1995	20	prose	Lexesp
a15	unknown	unknown	1978-1995	20	prose	Lexesp

ID	Title	Author	Date	Century	Genre	Collection
a8	unknown	unknown	1978-1995	20	prose	Lexesp
dc3	unknown	unknown	1978-1995	20	prose	Lexesp
e3	unknown	unknown	1978-1995	20	prose	Lexesp
n8	unknown	unknown	1978-1995	20	prose	Lexesp
t17	unknown	unknown	1978-1995	20	prose	Lexesp
t34.4	unknown	unknown	1978-1995	20	prose	Lexesp
t54	unknown	unknown	1978-1995	20	prose	Lexesp
a16	unknown	unknown	1978-1995	20	prose	Lexesp
a9	unknown	unknown	1978-1995	20	prose	Lexesp
dc4.1	unknown	unknown	1978-1995	20	prose	Lexesp
e4	unknown	unknown	1978-1995	20	prose	Lexesp
r1	unknown	unknown	1978-1995	20	prose	Lexesp
t18	unknown	unknown	1978-1995	20	prose	Lexesp
t35	unknown	unknown	1978-1995	20	newspaper	Lexesp
t55	unknown	unknown	1978-1995	20	newspaper	Lexesp
a17	unknown	unknown	1978-1995	20	newspaper	Lexesp
c1	unknown	unknown	1978-1995	20	newspaper	Lexesp
dc4.2	unknown	unknown	1978-1995	20	newspaper	Lexesp
e5	unknown	unknown	1978-1995	20	newspaper	Lexesp
r10.1a	unknown	unknown	1978-1995	20	newspaper	Lexesp
t19	unknown	unknown	1978-1995	20	newspaper	Lexesp
t36	unknown	unknown	1978-1995	20	newspaper	Lexesp
t56	unknown	unknown	1978-1995	20	newspaper	Lexesp
a18	unknown	unknown	1978-1995	20	newspaper	Lexesp
c101	unknown	unknown	1978-1995	20	newspaper	Lexesp
dc4.3	unknown	unknown	1978-1995	20	newspaper	Lexesp
e6	unknown	unknown	1978-1995	20	newspaper	Lexesp
r10.1b	unknown	unknown	1978-1995	20	newspaper	Lexesp
t2	unknown	unknown	1978-1995	20	newspaper	Lexesp
t37	unknown	unknown	1978-1995	20	newspaper	Lexesp
t57	unknown	unknown	1978-1995	20	newspaper	Lexesp

ID	Title	Author	Date	Century	Genre	Collection
a19	unknown	unknown	1978-1995	20	newspaper	Lexesp
c2	unknown	unknown	1978-1995	20	newspaper	Lexesp
dc5	unknown	unknown	1978-1995	20	newspaper	Lexesp
e7	unknown	unknown	1978-1995	20	newspaper	Lexesp
r10.1c	unknown	unknown	1978-1995	20	newspaper	Lexesp
t20	unknown	unknown	1978-1995	20	newspaper	Lexesp
t38	unknown	unknown	1978-1995	20	newspaper	Lexesp
t58	unknown	unknown	1978-1995	20	newspaper	Lexesp
a2	unknown	unknown	1978-1995	20	newspaper	Lexesp
c4	unknown	unknown	1978-1995	20	newspaper	Lexesp
dc6	unknown	unknown	1978-1995	20	newspaper	Lexesp
e8	unknown	unknown	1978-1995	20	newspaper	Lexesp
r10.2	unknown	unknown	1978-1995	20	newspaper	Lexesp
t21	unknown	unknown	1978-1995	20	newspaper	Lexesp
t39	unknown	unknown	1978-1995	20	newspaper	Lexesp
t59	unknown	unknown	1978-1995	20	newspaper	Lexesp
a20	unknown	unknown	1978-1995	20	newspaper	Lexesp
c5	unknown	unknown	1978-1995	20	newspaper	Lexesp
dc7	unknown	unknown	1978-1995	20	newspaper	Lexesp
e9	unknown	unknown	1978-1995	20	newspaper	Lexesp
r2	unknown	unknown	1978-1995	20	newspaper	Lexesp
t22	unknown	unknown	1978-1995	20	newspaper	Lexesp
t4	unknown	unknown	1978-1995	20	newspaper	Lexesp
t6	unknown	unknown	1978-1995	20	newspaper	Lexesp
a21	unknown	unknown	1978-1995	20	newspaper	Lexesp
c6	unknown	unknown	1978-1995	20	newspaper	Lexesp
dc8	unknown	unknown	1978-1995	20	newspaper	Lexesp
ed1	unknown	unknown	1978-1995	20	newspaper	Lexesp
r3	unknown	unknown	1978-1995	20	newspaper	Lexesp
t23	unknown	unknown	1978-1995	20	newspaper	Lexesp
t40	unknown	unknown	1978-1995	20	newspaper	Lexesp

ID	Title	Author	Date	Century	Genre	Collection
t60	unknown	unknown	1978-1995	20	newspaper	Lexesp
a22	unknown	unknown	1978-1995	20	newspaper	Lexesp
c7.1	unknown	unknown	1978-1995	20	newspaper	Lexesp
dc9	unknown	unknown	1978-1995	20	newspaper	Lexesp
ed2	unknown	unknown	1978-1995	20	newspaper	Lexesp
r4	unknown	unknown	1978-1995	20	newspaper	Lexesp
t24	unknown	unknown	1978-1995	20	newspaper	Lexesp
t41	unknown	unknown	1978-1995	20	newspaper	Lexesp
t61	unknown	unknown	1978-1995	20	newspaper	Lexesp
a23	unknown	unknown	1978-1995	20	newspaper	Lexesp
c7.2	unknown	unknown	1978-1995	20	newspaper	Lexesp
e1	unknown	unknown	1978-1995	20	newspaper	Lexesp
ed3	unknown	unknown	1978-1995	20	newspaper	Lexesp
r5	unknown	unknown	1978-1995	20	newspaper	Lexesp
t25	unknown	unknown	1978-1995	20	newspaper	Lexesp
t42	unknown	unknown	1978-1995	20	newspaper	Lexesp
t7	unknown	unknown	1978-1995	20	newspaper	Lexesp
a24	unknown	unknown	1978-1995	20	newspaper	Lexesp
c8	unknown	unknown	1978-1995	20	newspaper	Lexesp
e10	unknown	unknown	1978-1995	20	newspaper	Lexesp
ed4.1	unknown	unknown	1978-1995	20	didactics	Lexesp
r6	unknown	unknown	1978-1995	20	didactics	Lexesp
t26	unknown	unknown	1978-1995	20	didactics	Lexesp
t43	unknown	unknown	1978-1995	20	didactics	Lexesp
t8	unknown	unknown	1978-1995	20	didactics	Lexesp
a25	unknown	unknown	1978-1995	20	didactics	Lexesp
c9	unknown	unknown	1978-1995	20	didactics	Lexesp
e11	unknown	unknown	1978-1995	20	didactics	Lexesp
ed4.2	unknown	unknown	1978-1995	20	didactics	Lexesp
r7	unknown	unknown	1978-1995	20	didactics	Lexesp
t27	unknown	unknown	1978-1995	20	didactics	Lexesp

ID	Title	Author	Date	Century	Genre	Collection
t44	unknown	unknown	1978-1995	20	didactics	Lexesp
t9	unknown	unknown	1978-1995	20	didactics	Lexesp
a26	unknown	unknown	1978-1995	20	didactics	Lexesp
d1	unknown	unknown	1978-1995	20	didactics	Lexesp
e12	unknown	unknown	1978-1995	20	didactics	Lexesp
ed4.3	unknown	unknown	1978-1995	20	didactics	Lexesp
r8	unknown	unknown	1978-1995	20	didactics	Lexesp
t28	unknown	unknown	1978-1995	20	didactics	Lexesp
t45	unknown	unknown	1978-1995	20	didactics	Lexesp
a27	unknown	unknown	1978-1995	20	didactics	Lexesp
d2	unknown	unknown	1978-1995	20	didactics	Lexesp
e13	unknown	unknown	1978-1995	20	didactics	Lexesp
n1	unknown	unknown	1978-1995	20	didactics	Lexesp
r9	unknown	unknown	1978-1995	20	didactics	Lexesp
t29	unknown	unknown	1978-1995	20	didactics	Lexesp
t46	unknown	unknown	1978-1995	20	didactics	Lexesp
a28	unknown	unknown	1978-1995	20	didactics	Lexesp
d3	unknown	unknown	1978-1995	20	didactics	Lexesp
e14	unknown	unknown	1978-1995	20	didactics	Lexesp
n2.1	unknown	unknown	1978-1995	20	didactics	Lexesp
t1	unknown	unknown	1978-1995	20	didactics	Lexesp
t3	unknown	unknown	1978-1995	20	didactics	Lexesp
t47	unknown	unknown	1978-1995	20	didactics	Lexesp

Appendix B

FREQUENCY DATA

The table in the following pages lists the number of participial constructions per document ID, ordered by date of composition.

ID	Date	*estar* + PTCP	*haber* + PTCP	*ser* + PTCP	*tener* + PTCP
CID	1100	2	124	213	5
CZP	1214	81	397	826	77
SME	1215	1	23	73	3
VSM	1215	1	6	48	1
BER	1230	37	208	728	61
SMU	1230	5	28	179	0
TRO	1230	0	3	19	1
FGN	1237	4	229	972	29
ALX	1240	60	385	1070	71
FAR	1250	6	219	883	7
ALB	1250	2	72	291	0
GE5	1250	48	351	588	58
JUZ	1250	16	203	1996	10
LAP	1250	17	79	739	35
LEY	1250	12	290	1374	25
MOA	1250	23	231	413	14
PIC	1250	22	44	287	13
BDS	1250	8	118	322	9
IJ8	1250	86	679	2345	25
PFG	1250	20	88	302	17
POR	1250	8	24	191	3
REX	1250	0	1	22	2
TEO	1250	0	3	11	3
FRL	1255	4	64	839	10
CRZ	1259	1	21	351	3
ZRQ	1260	0	0	0	0
FJZ	1260	2	76	1568	15
LCB	1260	7	25	213	6
SPC	1260	15	470	2257	58
FN2	1270	8	211	748	30

ID	Date	*estar* + PTCP	*haber* + PTCP	*ser* + PTCP	*tener* + PTCP
FNV	1270	3	217	934	36
GCI	1270	168	3451	5933	91
GC3	1270	114	1547	1142	130
LAT	1270	50	563	1646	18
TFB	1270	2	9	49	6
TFC	1270	1	12	60	6
TFD	1270	0	12	54	5
TFE	1270	0	12	58	6
TFF	1270	0	11	52	4
TFG	1270	0	1	7	1
TFH	1270	1	9	66	8
TFI	1270	1	10	50	7
TFJ	1270	1	11	53	4
TFK	1270	1	11	48	3
TFL	1270	1	10	57	7
TFM	1270	3	11	57	6
TFN	1270	0	12	51	4
TFO	1270	0	9	54	3
TFP	1270	1	1	3	1
TFQ	1270	0	11	53	15
YMG	1276	1	3	229	1
TFA	1276	1	13	56	7
AST	1277	28	396	1796	12
RAB	1277	0	8	126	0
EE1	1280	117	615	1685	80
EE2	1280	167	950	1867	153
GE1	1280	161	1060	2229	48
GE4	1280	131	980	2660	87
CAS	1280	1	176	540	16
CAX	1280	39	283	328	53
CS4	1280	14	64	89	18
REY	1280	95	544	1106	128
FNA	1280	6	226	993	28
FNB	1280	7	210	770	31
FNC	1280	3	217	934	36
FND	1280	2	148	682	31
FNE	1280	3	210	917	25
FNF	1280	4	228	890	29
FNG	1280	6	243	868	39
FNH	1280	1	12	50	2
FNI	1280	1	175	674	31
FNJ	1280	4	196	685	25
FNK	1280	3	225	813	36
FNL	1280	3	200	953	34
FNM	1280	4	183	880	28
FNN	1280	3	193	898	30
FNO	1280	3	56	189	14
FNP	1280	7	208	1001	30

ID	Date	*estar* + PTCP	*haber* + PTCP	*ser* + PTCP	*tener* + PTCP
FNQ	1280	9	171	660	24
FNR	1280	4	211	988	34
FNS	1280	3	108	319	12
ACE	1283	21	33	343	7
VLT	1290	511	3621	4624	371
CC1	1293	5	15	98	6
CC2	1293	3	11	67	2
CD1	1293	48	144	431	34
CD2	1293	86	305	1039	51
ULT	1295	138	1117	1474	121
G2K	1300	84	673	1216	95
GE6	1300	1	20	78	1
ACR	1300	0	3	10	2
ARC	1300	1	0	2	0
CET	1300	1	50	93	5
DAN	1300	2	15	14	6
EMZ	1300	4	39	148	7
GLF	1300	0	1	2	1
GRF	1300	0	11	16	1
RCY	1300	0	4	17	1
SEM	1300	4	14	62	5
SMM	1300	0	4	16	0
LAB	1300	0	7	15	1
LA1	1300	1	3	8	0
CON	1300	0	1	5	0
REC	1300	3	22	57	2
PLA	1300	2	49	226	0
PA7	1300	19	287	1180	19
FYO	1307	3	195	396	3
AC1	1310	4	88	688	5
AC2	1310	3	80	615	8
AC3	1310	4	90	685	6
AC5	1310	5	88	707	6
BC1	1310	5	100	672	3
BC4	1310	5	101	681	5
AC6	1310	0	3	116	0
BRI	1313	3	86	1089	14
CF1	1315	11	68	988	14
CF2	1315	24	78	1131	21
CF3	1315	27	74	1164	35
LEO	1320	87	449	857	61
TRE	1320	4	67	147	4
MCR	1321	32	235	618	32
HCP	1325	1	17	53	1
MCE	1326	4	76	196	5
MCA	1330	10	43	153	8
MES	1330	18	378	805	38
FN1	1330	4	206	878	28

ID	Date	*estar* + PTCP	*haber* + PTCP	*ser* + PTCP	*tener* + PTCP
LUC	1331	32	243	383	32
MIN	1336	4	33	85	2
LBG	1340	27	57	219	19
LBS	1340	42	93	362	36
LBT	1340	13	34	99	7
LC1	1340	4	13	147	8
LC2	1340	2	18	143	6
LC3	1340	0	13	127	8
MAR	1340	1	23	39	1
MON	1340	21	113	543	8
MAS	1342	0	3	29	0
CRN	1344	80	671	1708	79
AO2	1348	5	36	512	6
FVA	1348	2	23	136	12
MRP	1350	1	104	187	5
NO1	1350	0	32	76	4
MEN	1350	9	74	342	5
RAH	1350	49	210	1139	62
VSC	1350	3	13	86	2
VIS	1350	3	15	521	0
MDR	1360	5	51	158	1
CQ1	1370	77	1323	1791	43
MOR	1370	8	273	524	0
ORS	1370	82	1042	2315	38
PL1	1370	23	650	1118	33
PL2	1370	23	592	1223	21
PL3	1370	15	506	1117	16
RAM	1370	17	369	930	21
SEC	1370	2	61	430	5
TRY	1370	41	330	617	16
TUC	1370	3	169	321	0
TUY	1370	6	508	1560	14
OA1	1370	5	38	515	5
TRS	1370	37	205	411	13
HJA	1370	25	117	749	42
EUT	1377	11	428	1529	9
EMP	1380	17	885	1141	31
MAN	1380	16	335	529	2
BUR	1380	1	173	537	16
PAL	1380	35	447	1184	10
CQ2	1385	114	2079	3074	68
ICR	1385	0	16	164	1
IHG	1385	0	17	30	1
LCA	1386	29	80	284	16
EUG	1390	26	213	667	35
G5R	1400	121	450	1009	85
AMZ	1400	32	276	424	24
APO	1400	13	113	261	15

ID	Date	*estar* + PTCP	*haber* + PTCP	*ser* + PTCP	*tener* + PTCP
ATA	1400	104	378	923	60
AUG	1400	35	180	388	4
BAE	1400	77	232	1589	117
BRA	1400	3	33	39	3
BRN	1400	43	139	1651	5
BSM	1400	1	5	18	2
CCC	1400	2	93	347	16
CCM	1400	5	50	181	5
CDS	1400	7	11	47	5
CGS	1400	56	175	514	41
CMN	1400	36	73	3161	54
CUZ	1400	65	238	1088	30
EOR	1400	13	103	385	17
ERG	1400	4	13	81	2
GYS	1400	5	31	158	14
NO2	1400	0	32	76	4
P13	1400	11	78	490	24
PN2	1400	18	57	157	11
PN5	1400	17	150	737	33
PN9	1400	3	36	312	5
PRS	1400	4	105	407	16
RHE	1400	14	28	213	3
DTM	1400	31	52	1437	25
LDA	1400	29	71	189	5
PSL	1400	3	21	156	2
ALV	1400	3	8	13	3
TAM	1410	208	302	585	56
SL2	1412	59	282	422	16
PER	1420	15	63	1398	31
OFI	1422	16	76	557	12
CIS	1423	20	23	250	23
LPA	1437	334	329	2506	97
CBO	1438	44	122	525	37
ARB	1441	6	580	942	27
DFN	1445	7	53	195	5
DIV	1445	49	50	249	24
DV2	1445	13	19	125	11
DON	1448	35	682	1180	38
AXI	1450	6	9	21	2
EVG	1450	2	9	15	5
JAM	1450	18	132	108	5
OLI	1450	23	146	152	17
SER	1450	6	8	77	2
LAA	1450	45	134	879	49
LVC	1450	64	78	549	32
TVV	1450	38	207	613	27
OVM	1452	37	159	1454	21
INV	1453	25	183	798	24

ID	Date	*estar* + PTCP	*haber* + PTCP	*ser* + PTCP	*tener* + PTCP
JPM	1453	0	17	25	0
LUN	1453	0	6	2	0
EIV	1454	1	3	2	0
MUS	1460	5	9	68	4
P10	1465	17	153	607	31
TA1	1465	4	39	199	5
TA2	1465	4	42	187	5
TAL	1465	1	6	46	0
MPM	1467	6	50	224	1
JCE	1470	1	57	65	0
TR1	1475	27	95	147	26
TR2	1475	17	73	87	20
VES	1477	11	41	228	17
G&G	1480	8	116	112	30
RRA	1480	112	187	1011	58
GOW	1483	0	3	82	1
DBM	1484	0	35	89	0
MTV	1484	95	419	3240	102
LTR	1485	3	34	58	8
MGO	1485	37	37	84	14
CLV	1486	35	112	307	43
APL	1488	5	66	139	2
CAR	1488	24	62	170	27
MEL	1489	15	630	1032	16
AYL	1490	19	59	78	28
CTY	1490	111	755	1320	75
IMI	1490	27	159	490	23
VAS	1490	5	133	209	6
Y88	1490	56	315	467	20
Y89	1490	54	309	445	19
SPO	1491	129	2343	10557	204
DHL	1492	7	145	49	4
GRM	1492	7	279	71	10
ABC	1493	83	318	613	28
CAP	1493	81	429	257	70
COL	1493	2	13	8	4
SL3	1493	55	259	442	13
SLC	1493	14	116	364	11
SLI	1493	9	189	161	16
SLM	1493	12	115	236	7
CAU	1493	16	122	6171	32
LIM	1494	60	997	606	34
CDP	1495	172	665	1792	102
DLH	1495	14	33	130	4
ENC	1496	119	260	547	105
BLA	1496	27	160	434	44
GLX	1498	31	13	152	17
LUX	1498	2	1	14	1

ID	Date	*estar* + PTCP	*haber* + PTCP	*ser* + PTCP	*tener* + PTCP
C01	1499	40	311	209	40
COR	1499	15	73	313	13
HER	1499	9	103	292	8
MAE	1499	5	4	5	0
OLV	1499	75	187	325	30
C02	1500	42	304	216	40
C07	1500	64	390	265	53
C08	1500	63	392	274	54
C17	1500	57	368	239	50
C21	1500	0	13	12	2
GOD	1500	0	1	4	0
HLC	1500	106	473	616	57
RHJ	1500	0	1	1	0
STR	1500	23	234	313	25
CHA	1500	3	44	52	6
CHS	1500	43	70	179	24
CIR	1500	6	18	617	6
CMY	1500	36	67	3101	44
ESP	1500	14	32	252	14
FIE	1500	21	120	730	10
FLO	1500	4	28	219	3
FOR	1500	8	9	131	6
GEN	1500	5	14	29	0
GIL	1500	15	33	854	12
GOM	1500	5	2	108	2
GOR	1500	248	114	3002	85
GRA	1500	19	258	4486	14
LOL	1500	2	5	190	1
MAC	1500	8	18	301	9
MLS	1500	2	16	1194	1
RES	1500	2	29	127	8
ROM	1500	0	0	23	1
SDM	1500	5	10	113	0
SEV	1500	49	65	928	11
SUM	1500	47	70	272	24
TED	1500	33	75	1000	5
TES	1500	3	41	262	5
TPH	1500	6	7	61	10
TRM	1500	12	57	589	5
TRP	1500	11	11	49	2
C03	1501	39	318	208	41
TRI	1501	59	712	767	36
GAH	1504	331	554	748	175
LTO	1505	9	38	99	3
C10	1507	62	377	265	50
C04	1511	68	400	278	57
C09	1512	62	350	256	54
CPC	1512	100	411	758	72

ID	Date	*estar* + PTCP	*haber* + PTCP	*ser* + PTCP	*tener* + PTCP
C05	1513	53	351	239	46
CHI	1513	12	78	391	22
C11	1514	64	396	275	54
BOT	1515	9	1	215	4
DDH	1517	1	1	2	1
LZ1	1517	30	118	61	20
LZ2	1517	31	115	50	18
LZ3	1517	28	108	52	17
ORT	1517	9	10	16	6
C06	1518	65	397	272	55
C12	1518	66	391	275	55
C13	1518	66	387	274	52
M18	1518	6	40	282	5
MAG	1519	15	210	140	2
M19	1519	8	37	277	6
P&V	1524	13	128	149	6
C14	1525	67	391	276	56
C15	1525	66	395	276	55
CEGH	1525	63	125	97	22
C16	1526	67	396	278	55
CDC	1526	34	192	188	21
C18	1528	65	394	276	55
C19	1529	66	389	276	59
C20	1540	65	402	287	54
CBN	1541	21	84	332	69
YSL	1542	92	305	911	77
GHOZ	1542	72	560	162	35
LAH	1551	237	419	202	70
GHIU	1554	33	127	63	22
GHLO	1554	108	506	333	65
LAG	1556	3	11	40	11
SDC	1567	66	217	84	93
TBC	1570	64	157	82	62
IND	1572	37	186	153	34
MRD	1580	61	477	145	95
PAR	1580	56	97	91	21
SAN	1583	10	23	11	4
TDM	1586	11	17	39	3
BGM	1589	27	69	72	7
DJC	1594	6	24	27	4
DPP	1594	3	17	13	0
BTP	1598	11	32	31	3
EXP	1598	21	121	66	22
MFF	1598	33	49	51	14
ALG	1599	66	47	48	13
ZAM	1599	9	33	17	4
LEB	1600	47	70	37	8
GHKZ	1604	530	3047	879	341

ID	Date	*estar* + PTCP	*haber* + PTCP	*ser* + PTCP	*tener* + PTCP
GHLN	1614	71	290	195	47
GHLM	1616	64	415	144	51
CEFT	1618	15	85	32	9
GHOT	1620	52	345	103	16
GHOV	1623	89	296	113	55
GHST	1626	57	308	47	40
GHOU	1627	26	121	22	5
GHIM	1628	35	176	35	7
CEFS	1636	38	177	51	5
CEFR	1651	38	155	24	4
GHNV	1691	29	151	156	11
GHQZ	1726	112	787	161	53
GIJS	1747	154	826	262	46
CEFQ	1773	53	571	98	16
GHOR	1782	12	55	16	9
CEFW	1800	18	77	34	4
CEFU	1806	27	259	21	9
GHIW	1818	80	231	139	14
GHNZ	1833	23	169	11	10
CEFZ	1835	44	160	31	13
GHNW	1836	24	97	13	1
GHOS	1836	113	807	205	49
GHLU	1837	12	35	7	0
GHOP	1837	8	115	16	4
GHIS	1842	33	241	83	13
CEGI	1844	23	164	25	6
GHOW	1846	46	171	49	11
GHIT	1848	9	74	27	4
GHIZ	1848	94	362	274	25
GHJK	1848	164	447	278	28
GHJL	1848	125	392	328	30
GHIV	1849	134	778	192	33
GHKU	1849	156	1033	140	47
GHIN	1854	65	392	56	7
GHMO	1858	569	4837	399	154
GHIO	1860	58	387	87	16
GHLS	1860	9	67	3	4
GHSW	1860	16	67	11	4
GHJM	1863	164	944	164	62
GHRT	1864	78	630	77	30
GHJU	1870	42	134	138	20
GHMZ	1870	215	1285	232	55
CEFP	1871	28	354	33	11
GHRV	1872	65	288	70	16
GHNT	1873	54	483	92	16
GHIL	1874	38	264	34	9
GHMP	1874	49	928	117	25
GHMR	1874	120	755	124	23

ID	Date	*estar* + PTCP	*haber* + PTCP	*ser* + PTCP	*tener* + PTCP
GIJK	1874	86	649	83	24
GHKW	1876	9	129	29	7
GHNS	1876	75	411	92	27
GHUV	1876	109	725	100	29
GHWZ	1876	67	571	107	24
GHMQ	1878	84	559	105	24
GHNO	1878	62	505	92	16
GHJQ	1880	26	382	208	31
GHIK	1881	30	240	13	8
GHMW	1881	184	1083	247	41
GHQW	1881	60	214	92	17
GHTW	1882	64	363	59	14
GIJQ	1882	53	163	75	15
GHPW	1883	142	734	113	56
GHTV	1883	66	403	96	15
GHLP	1884	387	2963	381	134
GHLZ	1884	12	27	8	1
GHMV	1884	112	575	95	32
GHNR	1884	126	684	117	35
GHPQ	1884	83	426	81	27
CEFV	1885	24	199	42	11
GHNU	1885	105	733	172	24
GHTU	1885	79	325	64	22
GHPZ	1886	161	913	175	67
GHQV	1886	73	399	100	52
GHIP	1887	70	258	118	29
GHIQ	1887	50	243	87	19
GHMU	1887	0	2	2	1
GIJO	1887	143	690	143	47
GHMN	1888	68	553	44	11
GHQS	1888	221	1115	162	55
GHRU	1888	115	839	184	50
GHSU	1888	31	202	11	8
CEFO	1889	63	492	82	12
GHPU	1889	192	1219	166	78
GHLQ	1890	100	926	124	42
GHLW	1890	150	818	179	72
GHQT	1890	217	1181	173	92
GHRS	1890	124	886	147	74
GHUW	1890	110	566	183	21
GIJN	1891	158	629	146	70
GHQU	1892	141	794	118	51
GHPS	1893	126	752	107	43
GHPT	1893	102	670	95	36
GHIJ	1894	117	154	178	24
GHJV	1894	110	545	135	42
GHPR	1894	76	417	62	30
GHRW	1894	128	1002	155	71

ID	Date	*estar* + PTCP	*haber* + PTCP	*ser* + PTCP	*tener* + PTCP
GHVW	1895	110	706	127	43
GIJM	1895	42	165	73	17
GIJR	1896	35	141	40	16
GHMS	1897	84	574	100	15
GHNQ	1897	57	616	124	23
GHTZ	1897	77	619	106	27
GHUZ	1897	72	550	100	25
GHKM	1898	59	314	59	23
GHLR	1898	102	936	161	47
GHVZ	1899	82	885	160	30
GHMT	1901	29	183	24	2
GIJP	1901	12	75	10	4
GHKN	1903	121	736	157	30
GHNP	1903	29	199	39	7
GHSZ	1903	88	640	74	27
GHJW	1904	74	634	94	16
GHKO	1904	96	717	120	21
GHPV	1904	24	121	18	13
GHKR	1905	69	746	113	26
GHQR	1906	172	1041	118	38
GIJL	1907	25	148	39	26
GHJT	1908	96	811	141	25
GHOQ	1912	29	287	66	18
GHJR	1913	187	1531	233	61
GHIR	1915	19	183	18	3
GHJS	1916	144	1359	182	27
GHKS	1916	39	254	37	9
GHJP	1918	22	177	58	8
GHJO	1919	71	309	38	14
GHJZ	1919	163	1282	183	33
GHKL	1919	92	722	175	27
GHLV	1919	30	213	72	13
GHJN	1920	158	601	113	47
GHKV	1920	49	467	59	10
GHRZ	1920	115	594	118	30
GHKQ	1921	53	649	79	26
GHKP	1922	104	850	128	30
GHKT	1922	89	975	102	36
GHLT	1922	56	269	27	33
GHSV	1922	18	107	21	5
a29	1978-1995	1	29	5	0
d4	1978-1995	8	352	39	1
e15	1978-1995	9	191	68	4
n2.2	1978-1995	0	33	2	1
t10	1978-1995	8	245	13	3
t30	1978-1995	7	142	16	5
t48	1978-1995	17	470	27	4
a1	1978-1995	5	211	14	2

ID	Date	*estar* + PTCP	*haber* + PTCP	*ser* + PTCP	*tener* + PTCP
a3	1978-1995	0	41	1	0
d5.1	1978-1995	1	32	6	0
e16	1978-1995	11	278	73	1
n2.3	1978-1995	21	320	35	4
t11	1978-1995	4	204	2	4
t31	1978-1995	8	159	13	0
t49	1978-1995	17	426	20	1
a10	1978-1995	7	187	15	2
a30	1978-1995	1	113	1	1
d5.2	1978-1995	0	5	2	0
e17	1978-1995	13	276	70	8
n3	1978-1995	20	492	69	6
t12	1978-1995	6	196	7	3
t32	1978-1995	10	176	14	1
t5	1978-1995	7	224	16	2
a11	1978-1995	4	247	24	2
a4	1978-1995	1	21	2	1
d5.3	1978-1995	15	348	42	4
e18.1	1978-1995	0	17	0	0
n4	1978-1995	23	511	55	2
t13	1978-1995	5	184	13	4
t33	1978-1995	5	240	19	4
t50	1978-1995	17	472	33	9
a12	1978-1995	4	117	2	1
a5	1978-1995	3	292	14	2
dc1	1978-1995	9	322	70	2
e18.2	1978-1995	0	19	1	0
n5	1978-1995	17	490	60	3
t14	1978-1995	6	208	11	1
t34.1	1978-1995	6	60	5	1
t51	1978-1995	6	313	45	1
a13	1978-1995	5	145	23	1
a6	1978-1995	7	257	23	3
dc10	1978-1995	19	284	61	2
e18.3	1978-1995	0	57	3	0
n6	1978-1995	15	533	54	7
t15	1978-1995	2	207	16	4
t34.2	1978-1995	1	107	7	0
t52	1978-1995	19	407	32	5
a14	1978-1995	0	87	8	0
a7	1978-1995	3	187	27	1
dc2	1978-1995	13	318	57	4
e2	1978-1995	3	131	21	1
n7	1978-1995	27	583	121	8
t16	1978-1995	2	190	14	3
t34.3	1978-1995	4	81	3	0
t53	1978-1995	16	481	24	7
a15	1978-1995	8	234	19	0

ID	Date	*estar* + PTCP	*haber* + PTCP	*ser* + PTCP	*tener* + PTCP
a8	1978-1995	5	110	11	0
dc3	1978-1995	10	281	54	2
e3	1978-1995	1	91	21	1
n8	1978-1995	24	574	105	6
t17	1978-1995	4	131	13	1
t34.4	1978-1995	2	39	2	0
t54	1978-1995	9	520	19	4
a16	1978-1995	7	130	9	2
a9	1978-1995	0	7	3	0
dc4.1	1978-1995	3	46	8	0
e4	1978-1995	1	189	17	1
r1	1978-1995	14	502	41	6
t18	1978-1995	4	256	5	1
t35	1978-1995	11	253	27	5
t55	1978-1995	14	513	24	4
a17	1978-1995	2	78	13	1
c1	1978-1995	10	175	5	4
dc4.2	1978-1995	0	28	1	0
e5	1978-1995	4	85	28	1
r10.1a	1978-1995	3	23	2	0
t19	1978-1995	3	182	17	0
t36	1978-1995	7	282	14	5
t56	1978-1995	13	586	20	2
a18	1978-1995	2	68	5	1
c101	1978-1995	6	163	6	3
dc4.3	1978-1995	8	216	30	3
e6	1978-1995	5	133	25	1
r10.1b	1978-1995	5	22	2	1
t2	1978-1995	4	70	7	1
t37	1978-1995	5	304	9	2
t57	1978-1995	12	468	18	3
a19	1978-1995	0	16	2	0
c2	1978-1995	3	110	20	1
dc5	1978-1995	18	246	67	5
e7	1978-1995	3	98	28	1
r10.1c	1978-1995	0	31	5	0
t20	1978-1995	5	223	9	2
t38	1978-1995	8	249	14	2
t58	1978-1995	20	406	17	5
a2	1978-1995	2	15	0	0
c4	1978-1995	2	92	8	1
dc6	1978-1995	9	288	77	0
e8	1978-1995	4	119	34	1
r10.2	1978-1995	8	344	44	2
t21	1978-1995	7	187	6	1
t39	1978-1995	6	240	22	4
t59	1978-1995	20	369	32	9
a20	1978-1995	0	51	3	0

ID	Date	*estar* + PTCP	*haber* + PTCP	*ser* + PTCP	*tener* + PTCP
c5	1978-1995	2	78	10	0
dc7	1978-1995	13	307	82	2
e9	1978-1995	3	154	35	3
r2	1978-1995	24	572	59	5
t22	1978-1995	6	188	18	4
t4	1978-1995	5	228	15	2
t6	1978-1995	3	241	8	1
a21	1978-1995	5	302	30	5
c6	1978-1995	1	198	10	4
dc8	1978-1995	10	318	51	2
ed1	1978-1995	10	627	37	5
r3	1978-1995	15	370	47	8
t23	1978-1995	3	253	9	1
t40	1978-1995	11	214	9	1
t60	1978-1995	11	340	20	3
a22	1978-1995	2	97	4	1
c7.1	1978-1995	0	4	2	0
dc9	1978-1995	15	252	75	4
ed2	1978-1995	10	463	59	3
r4	1978-1995	19	436	72	7
t24	1978-1995	2	205	7	3
t41	1978-1995	5	350	22	3
t61	1978-1995	13	482	27	3
a23	1978-1995	2	177	34	1
c7.2	1978-1995	0	31	3	0
e1	1978-1995	3	109	18	2
ed3	1978-1995	12	469	56	6
r5	1978-1995	13	439	59	3
t25	1978-1995	9	168	12	1
t42	1978-1995	10	449	12	6
t7	1978-1995	4	194	12	1
a24	1978-1995	4	80	24	0
c8	1978-1995	1	8	1	0
e10	1978-1995	4	161	44	1
ed4.1	1978-1995	0	24	4	0
r6	1978-1995	12	450	51	4
t26	1978-1995	11	208	12	6
t43	1978-1995	9	411	37	4
t8	1978-1995	2	162	12	2
a25	1978-1995	2	209	21	1
c9	1978-1995	3	91	12	0
e11	1978-1995	4	143	39	2
ed4.2	1978-1995	0	4	0	0
r7	1978-1995	8	458	56	4
t27	1978-1995	7	260	10	3
t44	1978-1995	24	381	17	4
t9	1978-1995	7	234	12	5
a26	1978-1995	9	282	26	1

ID	Date	*estar* + PTCP	*haber* + PTCP	*ser* + PTCP	*tener* + PTCP
d1	1978-1995	16	582	30	3
e12	1978-1995	5	130	38	2
ed4.3	1978-1995	6	315	31	3
r8	1978-1995	26	575	37	8
t28	1978-1995	6	229	22	4
t45	1978-1995	12	453	34	0
a27	1978-1995	3	54	2	1
d2	1978-1995	8	558	30	1
e13	1978-1995	7	156	31	4
n1	1978-1995	12	424	47	4
r9	1978-1995	21	475	63	3
t29	1978-1995	9	202	18	1
t46	1978-1995	16	475	23	5
a28	1978-1995	5	231	11	0
d3	1978-1995	10	581	39	4
e14	1978-1995	10	456	47	2
n2.1	1978-1995	2	83	6	0
t1	1978-1995	4	251	15	1
t3	1978-1995	2	231	17	3
t47	1978-1995	14	366	48	10

Appendix C

ASPECTUAL CLASSES OF VERBS

The number in parentheses for each verbs indicates the total number of tokens (finite and non-finite forms included) in the 20th century corpus. The total number of types for each class is indicated in parenthesis in the section headings.

C.1 Accomplishments (Types: 50)

arrasar 'raze' (70), *arreglar* 'fix' (468), *ascender* 'ascend' (267), *atravesar* 'cross' (611), *bajar* 'lower' (1484), *borrar* 'delete' (339), *cantar* 'sing' (1098), *cavar* 'dig' (46), *cocinar* 'cook' (67), *colocar* 'place' (912), *comer* 'eat' (1808), *concretar* 'specify' (85), *construir* 'build' (614), *convertir* 'convert' (2316), *crear* 'create' (878), *cruzar* 'cross' (799), *cubrir* 'cover' (1119), *descender* 'descend' (586), *destruir* 'destroy' (336), *dibujar* 'draw' (226), *doblar* 'fold' (276), *elaborar* 'develop' (195), *encajar* 'fit' (127), *escalar* 'scalar' (58), *esconder* 'hide' (376), *escribir* 'write' (2318), *fabricar* 'manufacture' (345), *grabar* 'record' (111), *huir* 'escape' (1174), *hundir* 'sink' (499), *interpretar* 'interpret' (365), *leer* 'read' (2063), *levantar* 'lift' (2270), *pelar* 'peel' (59), *pintar* 'paint' (445), *planchar* 'iron' (36), *preparar* 'prepare' (862), *producir* 'produce' (2500), *provocar* 'provoke' (985), *recitar* 'recite' (150), *recorrer* 'tour' (614), *recuperar* 'recover' (464), *regresar* 'return' (814), *reparar* 'repair' (248), *repasar* 'review' (134), *sacar* 'get' (1992), *secar* 'dry' (187), *subir* 'upload' (1634), *tejer* 'weave' (95), *transcribir* 'transcribe' (35)

C.2 Achievements (Types: 50)

abrir 'open' (3021), *acertar* 'hit' (266), *adquirir* 'acquire' (796), *alcanzar* 'achieve' (1395), *apagar* 'turn off' (350), *caer* 'fall' (2960), *capturar* 'capture' (85), *cerrar* 'close' (1650), *cesar* 'cease' (406), *comenzar* 'start' (2452), *conseguir* 'get' (2367), *convertir* 'convert' (2316), *dejar* 'leave' (6987), *descubrir* 'discover' (1583), *despertar* 'awake' (1127), *destruir* 'destroy' (336), *detectar* 'detect' (301), *empezar* 'start' (3129), *encender* 'turn on' (612), *encontrar* 'find' (5789), *enterar* 'find out' (678), *entrar* 'enter' (3519), *estallar* 'burst' (359), *finalizar* 'finalize' (117), *firmar* 'sign' (323), *franquear* 'go through' (42), *ganar* 'win' (1657), *heredar* 'inherit' (156), *iniciar* 'start' (773), *llegar* 'arrive' (8882), *lograr* 'achieve' (1497), *matar* 'kill' (1488), *morir* 'die'

(3116), *nacer* 'be born' (1292), *olvidar* 'forget' (1927), *parar* 'stop' (638), *partir* 'cut, leave' (661), *perder* 'lose' (3351), *recibir* 'receive' (2130), *reconocer* 'recognize' (1703), *recordar* 'remember' (2941), *romper* 'break' (1244), *salir* 'go out' (5927), *sentar* 'sit' (1878), *surgir* 'arise' (916), *terminar* 'end' (1876), *traspasar* 'transfer' (128), *tropezar* 'stumble' (328), *venir* 'come' (5424), *volver* 'return' (6507)

C.3 Activities (Types: 50)

acariciar 'stroke' (470), *andar* 'walk' (1742), *bailar* 'dance' (447), *buscar* 'search' (2539), *chirriar* 'squeak' (51), *circular* 'circulate' (308), *conducir* 'lead' (779), *conspirar* 'conspire' (26), *continuar* 'continue' (1403), *coquetear* 'flirt' (21), *correr* 'run' (2273), *crecer* 'grow' (856), *crujir* 'creak' (123), *dormir* 'sleep' (1549), *empujar* 'push' (511), *escuchar* 'listen' (1661), *escupir* 'spit' (107), *esperar* 'wait' (3364), *explorar* 'explore' (135), *fumar* 'smoke' (529), *golpear* 'hit' (404), *gritar* 'shout' (1102), *hablar* 'talk' (7362), *insultar* 'insult' (175), *jugar* 'play' (1484), *llorar* 'mourn' (1064), *llover* 'rain' (285), *mear* 'piss' (43), *mecer* 'rock' (72), *mirar* 'look' (5395), *nadar* 'swim' (175), *nevar* 'snow' (34), *observar* 'observe' (1108), *pasear* 'walk' (641), *pensar* 'think' (5762), *perseguir* 'pursue' (503), *reír* 'laugh' (1528), *rodar* 'roll' (340), *roncar* 'snore' (48), *rugir* 'roar' (170), *sonreír* 'smile' (1499), *soñar* 'dream' (544), *temblar* 'tremble' (626), *trabajar* 'work' (1850), *tronar* 'thunder' (46), *usar* 'use' (680), *utilizar* 'use' (1238), *viajar* 'travel' (505), *vibrar* 'vibrate' (103), *volar* 'fly' (463)

C.4 States (Types: 50)

abundar 'abound' (139), *admirar* 'admire' (632), *adorar* 'adore' (265), *agradar* 'please' (97), *amar* 'love' (1133), *anhelar* 'long for' (49), *atañer* 'appertain' (42), *bastar* 'be enough' (1058), *caber* 'fit' (712), *carecer* 'lacking' (514), *comprender* 'understand' (1677), *concernir* 'concern' (45), *confiar* 'trust' (462), *conocer* 'know' (4635), *consistir* 'consist' (667), *constar* 'record' (212), *costar* 'cost' (859), *creer* 'believe' (5881), *depender* 'depend' (520), *desear* 'wish' (1561), *detestar* 'detest' (93), *dudar* 'doubt' (595), *entender* 'understand' (2073), *entrañar* 'carry' (46), *envidiar* 'envy' (141), *equivaler* 'be equivalent' (161), *existir* 'exist' (2558), *gustar* 'like' (2038), *implicar* 'involve' (348), *incumbir* 'be incumbent' (15), *lamentar* 'regret' (294), *merecer* 'deserve' (700), *necesitar* 'need' (2052), *odiar* 'hate' (363), *parecer* 'seem' (9146), *permanecer* 'remain' (1506), *pertenecer* 'pertain' (746), *poseer* 'possess' (980), *preferir* 'prefer' (916), *querer* 'want' (10092), *repugnar* 'disgust' (57), *requerir* 'require' (414), *saber* 'know' (12211), *sentir* 'feel' (4409), *significar* 'mean' (770), *sobrar* 'be left' (214), *suponer* 'suppose' (1620), *temer* 'fear' (1181), *tener* 'have' (27995), *valer* 'be worth' (945)

Appendix D

OTHER CLASSES OF VERBS

The number in parentheses for each verbs indicates the total number of tokens (finite and non-finite forms included) in the 20th century corpus. The total number of types for each class is indicated in parenthesis in the section headings.

D.1 Degree achievements (100 verbs)

D.1.1 Simple (*congelar* 'freeze') (57 verbs)

acelerar 'accelerate' (167), *agriar* 'sour' (12), *amanecer* 'dawn' (139), *amargar* 'embitter' (74), *ampliar* 'enlarge' (177), *ascender* 'ascend' (267), *aumentar* 'increase' (783), *avanzar* 'advance' (1122), *bajar* 'come down' (1484), *calentar* 'heat' (166), *calmar* 'calm' (114), *congelar* 'freeze' (43), *corroer* 'corrode' (24), *crecer* 'grow' (856), *debilitar* 'weaken' (66), *decaer* 'decay' (45), *degradar* 'degrade' (38), *derretir* 'thaw' (40), *desarrollar* 'develop' (786), *descender* 'descend' (586), *deteriorar* 'deteriorate' (82), *dilatar* 'dilate' (114), *disminuir* 'decrease' (298), *entumecer* 'benumb' (24), *erosionar* 'erode' (24), *estirar* 'stretch' (156), *evaporar* 'evaporate' (40), *fermentar* 'ferment' (22), *fundir* 'melt' (142), *incrementar* 'increase' (129), *inflamar* 'inflame' (70), *madurar* 'mature' (46), *mejorar* 'improve' (356), *menguar* 'wane' (35), *oscurecer* 'obscure' (70), *oxidar* 'oxidize' (26), *progresar* 'progress' (75), *reducir* 'reduce' (667), *sanar* 'heal' (27), *secar* 'dry' (187), *subir* 'upload' (1634), *templar* 'temper' (43), *vaciar* 'empty' (181)

D.1.2 Derived (*alargar* 'lengthen') (43 verbs)

abaratar 'cheapen' (23), *ablandar* 'soften' (50), *abreviar* 'abbreviate' (29), *acercar* 'get close' (1328), *aclarar* 'clarify' (315), *acobardar* 'daunt' (24), *acomodar* 'accommodate' (159), *acortar* 'shorten' (37), *acrecentar* 'increase' (62), *adelgazar* 'get thin' (62), *afear* 'disfigure' (15), *afinar* 'tune' (33), *agrandar* 'enlarge' (130), *agravar* 'aggravate' (62), *ahondar* 'delve' (23), *alargar* 'lengthen' (187), *alejar* 'move eaway' (668), *aligerar* 'lighten' (28), *alisar* 'smooth' (27), *amansar* 'tame' (22), *aminorar* 'lessen' (26), *aplanar* 'flatten' (10), *aproximar* 'approximate' (426), *atenuar* 'attenuate' (39), *embellecer* 'beautify' (36), *embobar* 'amaze' (36), *emborrachar* 'get drunk' (62), *embravecer* 'enrage' (21), *embrutecer* 'coarsen' (20), *empeorar* 'worsen' (54), *empequeñecer* 'become smaller' (51) *empobrecer* 'impoverish' (21), *enajenar* 'alienate' (24), *encarecer* 'become more expensive' (33), *encorvar* 'bend' (82), *encrespar* 'curl' (34), *enderezar*

'straighten' (66), *endulzar* 'sweeten' (27), *endurecer* 'harden' (84), *enflaquecer* 'get thin' (15), *enfriar* 'cool' (59), *engordar* 'fatten' (76), *engrandecer* 'enlarge' (19), *engrosar* 'swell' (34), *enloquecer* 'madden' (131), *enmudecer* 'fall silent' (48), *ennegrecer* 'blacken' (57), *ennoblecer* 'ennoble' (21), *enrarecer* 'rarefy' (20), *enriquecer* 'enrich' (109), *enrojecer* 'blush' (157), *ensordecer* 'deafen' (10), *ensuciar* 'dirty' (50), *enternecer* 'soften' (65), *entorpecer* 'hinder' (34), *enturbiar* 'cloud' (40), *envejecer* 'age' (121)

D.2 Object experiencer psychological verbs (100 verbs)

abrumar 'overwhelm' (126), *aburrir* 'bore' (193), *acongojar* 'wring' (22), *acosar* 'harass' (127), *afectar* 'affect' (471), *afligir* 'afflict' (42), *agobiar* 'overwhelm' (43), *alentar* 'encourage' (78), *aliviar* 'alleviate' (153), *alterar* 'alter' (302), *alucinar* 'hallucinate' (15), *amedrentar* 'intimidate' (31), *amenazar* 'threaten' (408), *angustiar* 'distress' (53), *animar* 'encourage' (427), *anonadar* 'annihilate' (72), *apaciguar* 'appease' (37), *apasionar* 'have a passion' (97), *apenar* 'distress' (31), *apesadumbrar* 'apesadumbrar' (19), *apiadar* 'apiadar' (39), *arrepentir(se)* 'repent' (183), *arrepentirse* 'repent' (72), *asombrar* 'astonish' (286), *asustar* 'frighten' (365), *atemorizar* 'terrify' (43), *aterrar* 'terrify' (151), *aterrorizar* 'terrorize' (49), *atormentar* 'torment' (107), *aturdir* 'stun' (84), *aturdir* 'stun' (84), *avergonzar* 'shame' (221), *azorar* 'embarrass' (23), *cabrear* 'get angry' (20), *cohibir* 'restrain' (38), *compungir* 'compungir' (18), *confundir* 'confuse' (439), *conmocionar* 'shake up' (13), *consolar* 'comfort' (161), *consternar* 'dismay' (28), *contrariar* 'antagonize' (60), *decepcionar* 'disappoint' (61), *deprimir* 'depress' (29), *desencantar* 'disenchant' (14), *desesperar* 'despair' (211), *deshonrar* 'dishonor' (28), *desilusionar* 'disappoint' (20), *deslumbrar* 'dazzle' (99), *desmoralizar* 'demoralize' (18), *desolar* 'desolate' (84), *despistar* 'mislead' (25), *desquiciar* 'unhinge' (22), *devastar* 'devastate' (18), *distraer* 'distract' (233), *embelesar* 'ravish' (21), *emocionar* 'excite' (135), *empecinar* 'obstinates' (19), *enamorar* 'fall in love' (331), *encolerizar* 'anger' (21), *enfadar* 'get angry' (152), *enfurecer* 'infuriate' (64), *enojar* 'get angry' (43), *enorgullecer* 'make proud' (54), *entretener* 'entertain' (267), *entristecer* 'grieve' (62), *entusiasmar* 'get excited' (161), *espantar* 'scare' (127), *estimular* 'stimulate' (171), *exaltar* 'exalt' (91), *exasperar* 'exasperate' (68), *excitar* 'excite' (191), *extenuar* 'exhaust' (21), *fascinar* 'fascinate' (114), *fastidiar* 'annoy' (80), *frustrar* 'frustrate' (32), *honrar* 'honor' (65), *humillar* 'humiliate' (101), *ilusionar* 'excite' (42), *impacientar* 'irritate' (49), *impresionar* 'impress' (207), *indignar* 'be outraged' (253), *inspirar* 'inspire' (515), *interesar* 'interest' (773), *irritar* 'irritate' (206), *maravillar* 'wonder' (30), *molestar* 'disturb' (440), *mosquear* 'piss' (13), *motivar* 'motivate' (67), *obsesionar* 'obsess' (86), *ofender* 'offend' (170), *ofuscar* 'obfuscate' (14), *oprimir* 'oppress' (140), *perturbar* 'disturb' (109), *preocupar* 'worry' (633), *reconfortar* 'comfort' (26), *relajar* 'relax' (51), *satisfacer* 'meet' (480), *seducir* 'seduce' (132), *sorprender* 'surprise' (691), *tranquilizar* 'calm down' (188), *trastornar* 'unsettle' (122)

D.3 Incremental theme verbs (*comer* 'eat') (25 verbs)

almorzar 'have lunch' (153), *bailar* 'dance' (447), *beber* 'drink' (905), *caminar* 'walk' (754), *cantar* 'sing' (1098), *cenar* 'dine' (255), *comer* 'eat' (1808), *construir* 'build' (614), *consumir* 'consume' (319), *correr* 'run' (2273), *coser* 'sew' (160), *crear* 'create' (878), *desayunar* 'breakfast' (100), *dibujar* 'draw' (226), *diseñar* 'design' (120), *escribir* 'write' (2318), *fumar* 'smoke' (529), *jugar* 'play' (1484), *leer* 'read' (2063), *limpiar* 'clean' (299), *nadar* 'swim' (175), *pintar* 'paint' (445), *planchar* 'iron' (36), *pulir* 'polish' (42), *segar* 'mow' (24)

D.4 Extent predicates (*extender* 'extend') (20 verbs)

abarcar 'span' (187), *abarrotar* 'glut' (14), *atravesar* 'cross' (611), *cercar* 'fence' (66), *cubrir* 'cover' (1119), *desparramar* 'scatter' (40), *dispersar* 'disperse' (49), *distribuir* 'distribute' (173), *doblar* 'fold' (276), *ensanchar* 'widen' (57), *esparcir* 'spread' (290), *estrechar* 'narrow' (219), *expandir* 'expand' (32), *extender* 'extend' (689), *llenar* 'fill' (812), *ocupar* 'occupy' (1446), *ondular* 'wave' (36), *rodear* 'surround' (806), *salpicar* 'spatter' (67), *tapar* 'cover' (185)

D.5 Location and locatum verbs (*encarcelar* 'jail') (50 verbs)

abotonar 'button' (10), *acampar* 'camp' (36), *aclimatar* 'acclimatize' (10), *acorralar* 'corner' (30), *amordazar* 'gag' (13), *amueblar* 'furnish' (29), *aprisionar* 'imprison' (58), *arrinconar* 'corner' (24), *aterrizar* 'land' (52), *atrincherar* 'entrench' (12), *embalsamar* 'embalm' (11), *embarcar* 'ship' (157), *embocar* 'hole' (10), *embolsar* 'pocket' (12), *embozar* 'contain' (12), *empantanar* 'swamp' (10), *empapelar* 'paper' (16), *empaquetar* 'package' (18), *emplazar* 'site' (22), *empolvar* 'powder' (32), *encadenar* 'chain' (35), *encajar* 'fit' (127), *encaminar* 'route' (165), *encarcelar* 'imprison' (16), *encasquetar* 'pull on' (11), *encastillar* 'castle' (7), *encauzar* 'channel' (25), *enclavar* 'embed' (13), *encorsetar* 'constrict' (13), *encristalar* 'crystal' (10), *encuadrar* 'frame' (56), *encumbrar* 'elevate' (17), *enfilar* 'head' (47), *enfocar* 'focus' (66), *enfundar* 'sheathe' (40), *enganchar* 'hook' (80), *engarzar* 'hook' (20), *enguantar* 'cover with gloves' (16), *enjabonar* 'lather' (10), *enjaular* 'cage' (20), *enlazar* 'link' (85), *enmarcar* 'frame' (85), *enmascarar* 'mask' (33), *enraizar* 'root' (18), *enrejar* 'put in bars' (10), *ensangrentar* 'get stain with blood' (59), *ensillar* 'saddle' (13), *entablar* 'cover with planks' (89), *enterrar* 'bury' (259), *envenenar* 'poison' (101)

D.6 Other change of state and change of location verbs (result verbs) (50 verbs)

D.6.1 Change of state verbs (26 verbs)

abrir 'open' (3021), *activar* 'activate' (54), *anular* 'cancel' (139), *atar* 'tie' (196), *averiar* 'damage' (28), *caducar* 'expire' (10), *cambiar* 'change' (1693), *cerrar* 'close' (1650), *cortar* 'cut' (844), *deformar* 'deform' (43), *desnudar* 'undress' (97), *destrozar* 'shatter' (191), *destruir* 'destroy' (336), *eliminar* 'remove' (221), *extinguir* 'extinguish' (108), *manchar* 'stain' (101), *matar* 'kill' (1488), *morir* 'die' (3116), *nacer* 'be born' (1292), *pagar* 'pay' (1245), *partir* 'from' (661), *romper* 'break' (1244), *separar* 'separate' (663), *surgir* 'arise' (916), *transformar* 'transform' (521), *unir* 'unite' (1000), *vestir* 'dress' (947)

D.6.2 Change of location or motion verbs (24 verbs)

acceder 'gain access' (232), *acudir* 'come' (926), *alcanzar* 'reach' (1395), *allegar* 'arrive' (15), *archivar* 'file' (36), *armar* 'arm' (290), *arribar* 'arrive' (37), *caer* 'fall' (2960), *colocar* 'place' (912), *entrar* 'enter' (3519), *escapar* 'escape' (861), *fugar* 'escape' (11), *fugarse* 'escape' (33), *guardar* 'save' (1323), *huir* 'escape' (1174), *llegar* 'arrive' (8882), *meter* 'put' (1441), *pasar*

'pass' (8612), *provenir* 'come from' (172), *refugiar* 'shelter' (199), *regresar* 'return' (814), *salir* 'leave' (5927), *venir* 'come' (5424), *volver* 'return' (6507)

Appendix E

FREQUENCY OF AUXILIARIES WITH VERB CLASSES

	Accomplishments	Achievements	Activities	IL-states
12th c.	0	0	0	0
13th c.	193	142	15	15
14th c.	104	88	5	9
15th c.	217	305	25	37
16th c.	173	292	86	44
17th c.	58	100	11	20
18th c.	13	14	2	12
19th c.	424	808	79	124
20th c.	138	347	30	56

Table E.1: Frequency of *estar* with aspectual classes of verbs.

	Accomplishments	Achievements	Activities	IL-states
12th c.	3	23	6	1
13th c.	719	2831	607	538
14th c.	523	1781	418	468
15th c.	586	2184	457	725
16th c.	556	1866	560	901
17th c.	229	738	172	354
18th c.	108	281	57	138
19th c.	2041	7984	1846	3976
20th c.	2833	9719	1935	3933

Table E.2: Frequency of *haber* with aspectual classes of verbs.

	Accomplishments	Achievements	Activities	IL-states
12th c.	3	24	2	0
13th c.	2190	7209	411	4637
14th c.	1386	5372	366	2338
15th c.	1685	4792	485	4183
16th c.	685	2487	251	1134
17th c.	45	190	30	153
18th c.	15	65	4	48
19th c.	296	615	231	759
20th c.	202	420	255	350

Table E.3: Frequency of *ser* with aspectual classes of verbs.

	Accomplishments	Achievements	Activities	IL-states
12th c.	0	1	0	0
13th c.	80	157	13	27
14th c.	50	75	18	8
15th c.	120	272	42	100
16th c.	95	272	34	135
17th c.	33	46	11	23
18th c.	16	7	1	8
19th c.	199	238	53	136
20th c.	57	109	31	39

Table E.4: Frequency of *tener* with aspectual classes of verbs.

	DeAch	OE (simple)	OE (derived)	IncCh	Ext	Loc	ChSta	ChLoc
13th c.	28	66	2	77	125	10	128	160
14th c.	17	32	2	31	40	6	62	71
15th c.	46	91	25	65	112	27	236	154
16th c.	77	146	7	63	132	39	243	124
17th c.	10	52	13	34	20	7	64	31
18th c.	3	15	4	8	10	3	13	4
19th c.	104	521	398	150	209	86	477	146
20th c.	56	228	87	43	98	65	200	70

Table E.5: Frequency of *estar* with other classes of verbs. DeAch stands for degree achievements, OE (simple) for object experiencer psychological verbs which are morphologically simple such as *enfadar*, OE (derived) for object experiencer psychological verbs which are morphologically simple such as *asustar*, IncCh stands for incremental change verbs, Ext stands for extent predicates, Loc for locatives, and ChSta and ChLoc for other change and state and change of location verbs.

	DeAch	OE (simple)	OE (derived)	IncCh	Ext	Loc	ChSta	ChLoc
13th c.	869	760	60	1190	543	348	4297	3388
14th c.	472	577	113	578	267	449	3321	2397
15th c.	972	696	270	767	432	257	2861	2417
16th c.	926	319	38	561	138	27	1004	1598
17th c.	24	33	24	17	10	8	94	67
18th c.	11	5	1	6	1	2	18	14
19th c.	138	335	87	138	65	53	287	259
20th c.	84	121	40	83	43	59	181	149

Table E.6: Frequency of *ser* with other classes of verbs.

Bibliography

Anttila, Raimo (1972). *An introduction to historical and comparative linguistics*. Macmillan, New York.

Aranovich, Raúl (2003). The semantics of auxiliary selection in old spanish. *Studies in Language*, **27**, 1–37.

Aranovich, Raúl (2009). From esse to ser: diachronic mismatches in the selection of perfect auxiliaries. In *Studia Linguistica in honorem Mariae Manoliu*.

Aronoff, Mark (1976). *Word Formation in Generative Grammar*. Linguistic Inquiry Monograph 1, MIT Press.

Baayen, Harald (2001). *Word Frequency Distributions*. Kluwer Academic Publishers, Dordrecht.

Baayen, R. Harald (2008). *Analyzing Linguistic Data. A Practical Introduction to Statistics using R*. Cambridge University Press.

Bach, Emmon (1986). The algebra of events. *Linguistics and Philosophy*, **9**, 5–16.

Baron, Alistair (2011). *Dealing with Spelling Variation in Early Modern English Texts*. Ph.D. thesis, School of Computing and Communications, Lancaster University.

Baron, Alistair and Rayson, Paul (2008). Vard 2: A tool for dealing with spelling variation in historical corpora. In *Proceedings of the Postgraduate Conference in Corpus Linguistics*, Birmingham, UK. Aston University.

Bastardas-Parera, Juan (1953). *Particularidades sintácticas del latín medieval*. CSIC.

Batllori, Montserrat (2011). Diacronía de los verbos psicológicos: una propuesta de entrada léxica. In *Historia del léxico: perspectivas de investigación*. Iberoamericana / Vervuert, Madrid - Frankfurt.

Batllori, Montserrat and Roca, Francesc (to appear). Grammaticalization of ser and estar in romance. In *Grammatical Change: Origins, Nature, Outcomes*. Oxford University Press, Oxford.

Batllori, Montserrat; Castillo, M.E.; Urgell, F. Roca (2009). Relation between changes: the location and possessive grammaticalization path in spanish. In J. Rafel (ed.), *Diachronic Linguistics*. Girona (ESP): Servei de Publicacions de la UdG / Documenta Universitaria.

Bauer, Brigitte L.M. (2006). 'synthetic' vs. 'analytic' in romance: The importance of varieties. In R. Gess and D. Arteaga (eds.), *Historical Romance Linguistics. Retrospective and Perspectives*, pages 287–304. Benjamins, Amsterdam.

Beavers, John (in press). Resultative constructions. In R. I. Binnick (ed.), *The Oxford Handbook on Tense and Aspect*. Oxford University Press, Oxford.

Belvin, Robert and den Dikken, Marcel (1997). There, happens, to, be, have. *Lingua*, **101**, 151–183.

Benveniste, Emile (1968). Mutations of linguistic categories. In W. P. Lehmann and Y. Malkiel (eds.), *Directions for historical linguistics: A symposium*, pages 80–94. Texas University Press.

Benzing, Joseph (1931). Zur geschichte von ser als hilfszeitwort bei den intransitiven verben im spanischen. *Zeitschrift für romanische Philologie*, **51**, 385–460.

Biber, Douglas (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, **8**(4), 243–257.

Binnick, Robert (1991). *Time and the Verb*. Oxford University Press, Oxford.

Bird, Steven and Simons, Gary (2003). Seven dimensions of portability for language documentation and description. *Language*, **79**.

Birdsong, David (1999). Introduction: Whys and why nots of the period hypothesis for second language acquisition. In *Second language acquisition and the Critical Period Hypothesis*, pages 178–198. Lawrence Erlbaum Associates, Mahwah, NJ.

Bock, J. Kathryn and Kroch, Anthony S. (1989). The isolability of syntactic processing. In G. N. Carlson and M. K. Tanenhaus (eds.), *Linguistic structure in Language Processing*, pages 157–196. Kluwer, Dordrecht.

Bouzet, Jean (1953). Orígenes del empleo de estar: ensayo de sintaxis histórica. In *Estudios dedicados a Menéndez Pidal*, pages 1–58. CSIC.

Boyland, Joyce Tang (1996). *Morphosyntactic change in progress: A psycholinguistic approach*. Phd, University of California, Berkeley.

Brants, Thorsten (2000). Tnt - a statistical part-of-speech tagger. In *Proceedings of the Sixth Applied Natural Language Processing Conference ANLP-2000*, Seattle, WA.

Buck, R. A. (1993). Affectedness and other semantic properties of english denominal locative verbs. *American Speech*, **68**, 139–160.

Burzio, Luigi (1986). *Italian syntax: A government-binding approach*. Foris, Dordrecht.

Bybee, Joan (2003). Mechanisms of change in grammaticization: The role of frequency. In B. D. Joseph and J. Janda (eds.), *The Handbook of Historical Linguistics*, pages 602–623. Oxford: Blackwell.

Bybee, Joan (2010). Markedness: iconicity, economy and frequency. In J. J. Song (ed.), *Handbook of Linguistic Typology*. Oxford University Press.

Bybee, Joan; Pagliuca, W.; Perkins, R. (1991). Back to the future. In E. Traugott and B. Heine (eds.), *Approaches to Grammaticalization*, volume 2, pages 17–58. John Benjamins.

Bybee, Joan; Perkins, Revere; Pagliucca, William (1994). *The Evolution of Grammar: Tense, aspect, and modality in the languages of the world*. University of Chicago Press.

Bybee, Joan L. (1985). *Morphology: a Study of the Relation between Meaning and Form*. Number 9 in Typological Studies in Language. Benjamins.

Bybee, Joan L. and Dahl, Osten (1989). The creation of tense and aspect systems in the languages of the world. *Studies in Language*, **13**, 51–103.

Cann, Ronnie (2006). Semantic underspecification and the pragmatic interpretation of be. In K. von Heusinger and K. Turner (eds.), *Where Semantics Meets Pragmatics*. Elsevier.

Carlson, R. (1991). Grammaticalization of postpositions and word order in senufo languages. In E. Traugott and B. Heine (eds.), *Approaches to Grammaticalization*, volume 1, pages 201–224. John Benjamins.

Cennamo, Michela (2008). The rise and development of analytic perfects in italo-romance. In T. Eythórsson (ed.), *Grammatical Change and Linguistic Theory: the Rosendal Papers*. Amsterdam: Benjamins.

Chater, Nick; Tenenbaum, Joshua B.; Yuille, Alan (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences*, **10**, 287–291.

Church, Kenneth (2000). Empirical estimates of adaptation: The chance of two noriega's is closer to p/2 than p2. In *Proceedings of Coling*, pages 173–179.

Clark, E.V. and Clark, H.H. (1979). When nouns surface as verbs. *Language*, **55**, 767–811.

Cleveland, William S. (1981). Lowess: A program for smoothing scatterplots by robust locally weighted regression. *The American Statistician*, **35**.

Comrie, Bernard (1976). *Aspect. An Introduction to the Study of Verbal Aspect and Related Problems*. Cambridge University Press, Cambridge/ New York.

Comrie, Bernard (1985). *Tense*. Cambridge University Press, Cambridge / New York.

Copple, Mary T. (2009). *A Diachronic Study of the Spanish Perfect(ive): Tracking the Constraints on a Grammaticalizing Construction*. Ph.D. thesis, University of New Mexico.

Copple, Mary T. (2011). Tracking the constraints on a grammaticalizing perfect(ive). *Language Variation and Change*, **23**, 163–191.

Corfis, Ivy A.; O'Neill, John; Beardsley, Jr. Theodore S. (eds.) (1997). *Early Celestina Electronic Texts and Concordances*. Hispanic Seminary of Medieval Studies, Ltd. Madison, Wisconsin.

Cowper, Elizabeth (1989). Perfective -en is passive -en. In *Proceedings of the Eighth West Coast Conference on Formal Linguistics*, Stanford. CSLI.

Dahl, Osten (1985). *Tense and Aspect Systems*. Blackwell, Oxford.

de Miguel, Elena (1999). El aspecto léxico. In I. Bosque and V. Demonte (eds.), *Gramática descriptiva de la lengua española*, page 2971–3060. Espasa Calpe, Madrid.

Devitt, Dan (1990). The diachronic development of semantics in copulas. In *Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistics Society*, pages 103–115.

Dipper, Stefanie (2010). Pos-tagging of historical language data: First experiments. In *Semantic Approaches in Natural Language Processing: Proceedings of the Conference on Natural Language Processing (KONVENS 2010)*.

Dowty, David (1979). *Word meaning and Montague grammar: the semantics of verbs and times in generative semantics*. Springer.

Dowty, David (1991). Thematic proto-roles and argument selection. *Language*, **67**, 547–619.

Déchaine, Rose-Marie; Hoekstra, Teun; Rooryck, Johan (1994). Augmented and non-augmented have. In L. Nash and G. Tsoulas (eds.), *Proceedings of Langues et Grammaire*, pages 85–101.

Eckardt, Regine (2006). *Meaning change in grammaticalization: an enquiry into semantic reanalysis*. Oxford University Press.

Ellegard, Alvar (1953). The auxiliary do: The establishment and regulation of its use in english. In F. Behre (ed.), *Gothenburg Studies in English*. Almqvist and Wiksell, Stockholm.

Escandell-Vidal, Victoria and Leonetti, Manuel (2002). Coercion and the stage/individual distinction. In J. Gutierrez-Rexach (ed.), *From words to discourse*, page 159–179. Elsevier.

Evert, Stefan (2006). How random is a corpus? the library metaphor. *Zeitschrift für Anglistik und Amerikanistik*, **54**, 177 – 190.

Fischer, Olga (2010). An iconic analogical approach to grammaticalization. In J. Conradie, R. Johl, M. Beukes, O. Fischer, and C. Ljungberg (eds.), *Signergy. Iconicity in Language and Literature*, pages 279–298. Benjamins.

Fontana, Josep Maria (1993). *Phrase Structure and the Syntax of Clitics in the History of Spanish*. Ph.D. thesis, University of Pennsylvania.

Frank E. Harrell, Jr. (2001). *Regression Modeling Strategies. With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer.

Fridén, Georg (1948). *Studies on the tenses of the English verb from Chaucer to Shakespeare, with special reference to the late sixteenth century*. Harvard University Press, Cambridge, MA.

Gawron, Jean Mark (2005). Generalized paths. In *Proceedings of SALT XV*.

Gehrke, Berit (2012). Passive states. In V. Demonte and L. McNally (eds.), *Telicity, Change, and State: A Cross-Categorial View of Event Structure*. Clarendon Press.

Gehrke, Berit (to appear). Puzzled by adjectival passives. In R. Folli, C. Sevdali, and R. Truswell (eds.), *On Linguistic Interfaces II*. Oxford University Press, Oxford.

Gehrke, Berit and Sánchez-Marco, Cristina (2012). The role of by-phrases in adjectival passives. In *Workshop on Aspect and Argument Structure of Adjectives and Participles*, University of Greenwich.

Givón, Tom (1985). Function, structure, and language acquisition. In D. Slobin (ed.), *The crosslinguistic study of language acquisition*, volume 2, pages 1005–1028. Lawrence Erlbaum, Hillsdale, NJ.

Goldwater, Micah B.; Tomlinson, Marc T.; Echols, Catharine H.; Love, Bradley C. (2011). Structural priming as structure-mapping: Children use analogies from previous utterances to guide sentence production. *Cognitive Science*, **35**, 156–170.

Gopnik, A.; Glymour, C.; Sobel, D.; Schulz, L.; Kushnir, T.; Danks, D. (2004). A theory of causal learning in children: Causal maps and bayes nets. *Psychological Review*, **111**, 1–31.

Gries, Stefan Th. (2006). Exploring variability within and between corpora: some methodological considerations. *Corpora*, **1**(2), 109–151.

Gries, Stefan Th. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics*, **13**(4), 403–437.

Guéron, Jacqueline (1986). Le verbe 'avoir'. *Recherches Linguistiques*, **14**.

Haiman, John (1994). Ritualization and the development of language. In W. Pagliuca (ed.), *Perspectives on grammaticalization*, pages 3–28. John Benjamins, Amsterdam.

Han, Chung Hye and Kroch, Anthony (2000). The rise of do-support in english: Implications for clause structure. In *Proceedings of the North East Linguistic Society (NELS 30)*, Washington D.C. Georgetown University.

Hanssen, Federico (1912). La pasiva castellana. *Anales Universidad de Chile*, **CXXX**, 97–112.

Harley, Heidi (1998). You're having me on: Aspects of have. In J. Guéron and A. Zribi-Hertz (eds.), *La grammaire de la possession*. University Paris X – Nanterre, Paris.

Harre, Catherine E. (1991). *Tener + Past Participle. A case study in linguistic description*. Routledge, London, New York.

Harris, Alice C. and Campbell, Lyle (1995). *Historical Syntax in Cross-linguistic perspective*. Cambridge University Press.

Hay, Jen; Kennedy, Christopher; Levin, Beth (1999). Scale structure underlies telicity in 'degree achievements'. In T. Matthews and D. Strolovitch (eds.), *Semantics and Linguistic Theory 9*, page 127–144. CLC Publications, Ithaca, NY.

Heine, Bernd (1997). *Possession. Cognitive Sources, Forces and Grammaticalization*. Cambridge University Press.

Heine, Bernd and Kuteva, Tania (2002). *World lexicon of grammaticalization*. Cambridge University Press.

Heine, Bernd; Claudi, Ulrike; Hünnemeyer, Friederike (1991). *Grammaticalization: a Conceptual Framework*. University of Chicago Press, Chicago.

Herrera, María Teresa and de Fauve, María Estela González (eds.) (1997). *Concordancias electrónicas del corpus médico español*. Hispanic Seminary of Medieval Studies, Ltd. Madison, Wisconsin.

Hook, Peter E. (1991). The emergence of perfective aspect in indo-aryan languages. In E. C. Traugott and B. Heine (eds.), *Approaches to Grammaticalization*, volume 2, pages 59–89. Benjamins, Amsterdam.

Hopper, Paul J. and Traugott, Elizabeth Closs (2003). *Grammaticalization*. Cambridge University Press, 2nd edition.

Hudson Kam, Carla L. and Newport, Elissa L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, **1**, 151–195.

Iatridou, Sabine; Anagnostopoulou, Elena; Pancheva, Roumyana (2003). Observations about the form and meaning of the perfect. In A. Alexiadou and A. von Stechow (eds.), *Perfect Explorations*, pages 153–204. Mouton de Gruyter, Berlin.

Ide, Nancy and Romary, Laurent (2004). International standard for a linguistic annotation framework. *Natural Language Engineering*, **3**, 211–225.

Ide, Nancy and Romary, Laurent (2007). Towards international standards for language resources. In L. Dybkjaer, H. Hemsen, and W. Minker (eds.), *Evaluation of Text and Speech Systems*, pages 263–84. Springer.

Johannisson, Ture (1958). On the be and have constructions with mutative verbs. *Studia Linguistica*, **12**, 106–118.

Johnson, Jacqueline S.; Shenkman, Kenneth D.; Newport, Elissa L.; Medin, Douglas L. (1996). Indeterminacy in the grammar of adult language learners. *Journal of Memory and Language*, **35**, 335–352.

Joseph, Brian D. (1983). *The Synchrony and Diachrony of the Balkan Infinitive. A Study in Areal, General, and Historical Linguistics*. Cambridge University Press, Cambridge.

Joseph, Brian D. (2001). Is there such a thing as grammaticalization? *Language Sciences*, **23**, 163–186.

Jäger, Gerhard and Rosenbach, Anette (2008). Priming and unidirectional change. *Theoretical Linguistics*, **34**(2), 85–113.

Kasten, Llyod; Nitti, John; Jonxis-Henkemens, Wilhemina (eds.) (1997). *The Electronic Texts and Concordances of the Prose Works of Alfonso X, El Sabio*. Hispanic Seminary of Medieval Studies, Ltd. Madison, Wisconsin.

Kasten, Lloyd; Nitti, John; Jonxis-Henkemans, Wilhelmina; editors: Pedro Sánchez-Prieto Borja, Francisco Gago Jover. Copy; Díaz Moreno, Rocío; Trujillo Belso, Elena; Gago Jover, Francisco (eds.) (2011). *Obra en prosa de Alfonso X el Sabio*. University of Madison, Hispanic Seminary of Medieval Studies, Biblioteca Digital de Textos del Español Antiguo.

Kessler, Brett; Numberg, Geoffrey; Schütze, Hinrich (1997). Automatic detection of text genre. In *EACL '97 Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics*, PA, USA. Association for Computational Linguistics Stroudsburg.

Kilgarriff, Adam (2005). Language is never ever ever random. *Corpus Linguistics and Linguistic Theory*, **1**(2), 263–276.

Kratzer, Angelika (2000). Building statives.

Kratzer, Angelika (2004). Building resultatives. University of Amherst.

Krifka, Manfred (1989). Nominal reference, temporal constitution and quantification in event semantics. In J. van Benthem, R. Bartsch, and P. van Emde Boas (eds.), *Semantics and Contextual Expression*, pages 75–115. Foris, Dordrecht.

Kroch, Anthony (1989a). The loss of the verb-second constraint in middle english and middle french. In *Paper presented at the 18th NWAV Conference, Université de Montréal*.

Kroch, Anthony (1989b). Reflexes of grammar in patterns of language change. *Language Variation and Change*, **1**, 199–244.

Kroch, Anthony (1994). Morphosyntactic variation. In *Papers from the 30th Regional Meeting of the Chicago Linguistics Society: Parasession on Variation and Linguistic Theory*, pages 180–201.

Kroch, Anthony (2005). Modeling language change and language acquisition. University of Pennsylvania.

Kytö, M. (1991). *Manual to the Diachronic Part of the Helsinki Corpus of English Texts*. Department of English, University of Helsinki.

Kytö, Merja (1997). Be/have + past participle: The choice of the auxiliary with intransitives from late middle to modern english. In M. Rissanen and M. Kytö (eds.), *English in transition: Corpus-based studies in linguistic variation and genre styles*, page 17–86. Mouton de Gruyter, Berlin.

Labov, William (1981). Resolving the neogrammarian controversy. *Language*, **57**(2), 267–308.

Langacker, Ronald W. (1977). Syntactic reanalysis. In *Mechanisms of Syntactic Change*, pages 57–139. University of Texas Press, Austin.

Leech, Geoffrey (1971). *Meaning and the English verb*. Longman, London.

Lehmann, Christian (1985). Grammaticalization: synchronic variation and diachronic change. *Lingua e stile*, **20**, 303–318.

Levin, Beth (1993). *English Verb Classes and Alternations. A Preliminary Investigation*. The University of Chicago Press.

Levin, Beth and Rappaport, Malka (1994). A preliminary analysis of causative verbs in english. *Lingua*, **92**, 35–77.

Levin, Beth and Rappaport, Malka (1995). *Unaccusativity: At the syntax-semantics interface*. MIT Press, Cambridge, MA.

Lim, Chul; Lee, Kong; Kim, Gil (2005). Automatic genre detection of web documents. In *Natural Language Processing – IJCNLP 2004*, pages 310–319. Springer Berlin / Heidelberg.

Lord, C. (1973). Serial verbs in transition. *Studies in African Linguistics*, **4**, 269–296.

Luka, B. and Barsalou, L. W. (2005). Structural facilitation: mere exposure effects of grammatical acceptability as evidence for syntactic priming in comprehension. *Journal of Memory and Language*, **52**, 436–459.

Marín, Rafael (2000). *El componente aspectual de la predicación*. Ph.D. thesis, Universitat Autònoma de Barcelona.

Marín, Rafael (2004). *Entre ser y estar*. Arco Libros.

Marín, Rafael (2010). Spanish adjectives within bounds. In P. Cabredo and O. Matushansky (eds.), *Adjectives: Formal analyses in syntax and semantics*, pages 307–331. John Benjamins.

McCawley, James (1971). Tense and time reference in english. In C. Fillmore and D. Langendoen (eds.), *Studies in Linguistic Semantics*, pages 96–113. Holt, Rinehart and Winston, New York.

McCawley, James D. (1981). Notes on the english present perfect. *Australian Journal of Linguistics*, page 81–90.

McCoard, Robert W. (1978). *The English Perfect: Tense Choice and Pragmatic Inferences*. North-Holland Press, Amsterdam.

McFadden, Thomas and Alexiadou, Artemis (2005). Counterfactuals and be in the history of english. In J. Alderete and A. K. Chung-hye Han (eds.), *WCCFL 24: Proceedings of the 24th West Coast Conference on Formal Linguistics*, pages 272–280, Somerville, MA. Cascadilla Proceedings Project.

McFadden, Thomas and Alexiadou, Artemis (2010). Perfects, resultatives, and auxiliaries in earlier english. *Linguistic Inquiry*, **41**, 389–425.

McIntyre, Andrew (2006). The interpretation of german datives and english have. In D. Hole, A. Meinunger, and W. Abraham (eds.), *Datives and Other Cases*, pages 185–211. Benjamins.

Meillet, Antoine (1912). L'évolution des formes grammaticales. *Scientia (Rivista di Scienza)*, **12**(26). Reprinted in Meillet 1958: 130-148.

Mendeloff, Henry (1964). The passive voice in old spanish. *Romanistisches Jahrbuch*, page 269–287.

Michaelis, Laura A. (1994). The ambiguity of the english present perfect. *Journal of Linguistics*, **30**, 111–158.

Mittwoch, Anita (1988). Aspects of english aspect: On the interaction of perfect, progressive, and durational phrases. *Linguistics and Philosophy*, **11**, 203–254.

Mittwoch, Anita (2008). The english resultative perfect and its relationship to the experiential perfect and the simple past tense. *Linguistics and Philosophy*, **31**, 323–351.

Musan, Renate (2002). *The German Perfect. Its Semantic Composition and its Interaction with Temporal Adverbials*. Kluwer, Dordrecht, Boston and London.

Nebrija, Antonio (1517). *Reglas de orthographía en la lengua castellana*. Instituto Caro y Cuervo, Bogotá, 1977.

Newmeyer, Frederick J. (2001). Deconstructing grammaticalization. *Language Sciences*, **23**, 187–229.

Nitti, John and Kasten, Llyod (eds.) (1997). *The Electronic Texts and Concordances of Medieval Navarro-Aragonese Manuscripts*. Hispanic Seminary of Medieval Studies, Ltd. Madison, Wisconsin.

Oaksford, Mike and Chater, Nick (2007). *Bayesian Rationality: The probabilistic approach to human reasoning*. Oxford University Press.

O'Neill, John (ed.) (1999). *Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings*. Hispanic Seminary of Medieval Studies, Ltd. Madison, Wisconsin.

Padró, Lluís (2011). Analizadores multilingües en freeling. *Linguamatica*, **3**(2), 13–20.

Padró, Lluís and Stanilovsky, Evgeny (2012). Freeling 3.0: Towards wider multilinguality. In *Proceedings of the Language Resources and Evaluation Conference (LREC 2012), ELRA*, Istanbul, Turkey.

Padró, Lluís; Collado, Miquel; Reese, Samuel; Lloberes, Marina; Castellón, Irene (2010). Freeling 2.1: Five years of open-source language processing tools. In *Proceedings of 7th Language Resources and Evaluation Conference (LREC 2010)*, La Valletta, Malta. ELRA.

Pancheva, Roumyana (2003). The aspectual makeup of perfect participles and the interpretation of the perfect. In A. Alexiadou, M. Rathert, and A. von Stechow (eds.), *Perfect explorations*, page 277–306. Mouton, Berlin.

Perlmutter, David (1978). *Impersonal passives and the unaccusative hypothesis*. Berkeley Linguistic Society.

Perlmutter, David (1989). Multiattachment and the unaccusative hypothesis: The perfect auxiliary in italian. *Probus*, **1**, 63–119.

Petrenz, Philipp and Webber, Bonnie (2012). Robust cross-lingual genre classification through comparable corpora. In *The 5th Workshop on Building and Using Comparable Corpora*, pages 1–9.

Pickering, Martin J. and Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, **27**, 169–190.

Piñón, Christopher (2005). Adverbs of completion in an event semantics. In H. J. Verkuyl, H. de Swart, and A. van Hout (eds.), *Perspectives on Aspect*, pages 149–166. Springer.

Portner, Paul (2003). The temporal semantics and modal pragmatics of the perfect. *Linguistics and Philosophy*, **26**, 459–510.

Pountain, Christopher J. (1985). Copulas, verbs of possession and auxiliaries in old spanish: The evidence fro structurally interdependent changes. *Bulletin of Hispanic Studies*, **62**, 337–355.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rapp, Irene (1996). Zustand? passiv? — Überlegungen zum sogenannten zustandspassiv". *Zeitschrift fur Sprachwissenschaft*, **15.2**, 231–265.

Rappaport-Hovav, Malka and Levin, Beth (2001). An event structure account of english resultatives. *Language*, **4**, 766–797.

Rayson, Paul; Archer, D.; Baron, A.; Smith, N. (2007). Tagging historical corpora - the problem of spelling variation. In *Proceedings of Digital Historical Corpora, Dagstuhl-Seminar 06491, International Conference and Research Center for Computer Science*, Schloss Dagstuhl, Wadern, Germany.

Reali, Florencia and Griffiths, Thomas L. (2009). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition*, **11**, 317–328.

Rehm, Georg; Schonefeld, Oliver; Witt, Andreas; Hinrichs, Erhard; Reis, Marga (2009). Sustainability of annotated resources in linguistics: A web-platform for exploring, querying, and distributing linguistic corpora and other resources. *Literary and Linguistic Computing*, **24**(2), 193–210.

Remus, Robert and Bank, Mathias (2012). Textual characteristics of different-sized corpora. In *The 5th Workshop on Building and Using Comparable Corpora*, pages 148–, Istanbul, Turkey. LREC2012 Workshop.

Ritter, Elizabeth and Rosen, Sara (1997). The function of have. *Lingua*, **101**, 295–321.

Roberts, Ian and Roussou, Anna (2003). *Syntactic Change: A Minimalist Approach to Grammaticalization*. Cambridge University Press.

Rögnvaldsson, Eiríkur and Helgadóttir, Sigrún (2008). Morphological tagging of Old Norse texts and its use in studying syntactic variation and change. In *2nd Workshop on Language Technology for Cultural Heritage Data, LREC 2008 workshop*, Marrakech.

Romani, Patrizia (2006). Tiempos de formación romance i. los tiempos compuestos. In C. C. Company (ed.), *Síntaxis histórica de la lengua española. Primera parte: la frase verbal.*, pages 243–346. Universidad Nacional Autónoma de México, Fondo de Cultura Económica.

Rosenbach, Anette and Jäger, Gerhard (2008). Priming and unidirectional language change. *Theoretical Linguistics*, **34**(2), 85–113.

Rydén, Mats and Brorström, Sverker (1987). *The be/have variation with intransitives in English*. Almqvist & Wiksell, Stockholm.

Sagi, Eyal; Kaufmann, Stefan; Clark, Brady (2009). Semantic density analysis: Comparing word meaning across time and phonetic space. In R. Basili and M. Pennacchiotti (eds.), *Proceedings of the EACL 2009 Workshop on GEMS: GEometrical Models of Natural Language Semantics*, Athens.

Sanchez, María Nieves; Herrera, María Teresa; Zabía, María Purificación (eds.) (2003). *Textos medievales misceláneos*. Hispanic Seminary of Medieval Studies, Ltd. Madison, Wisconsin.

Schlücker, Barbara (2005). Event-related modifiers in German adjectival passives. In *Proceedings of Sinn und Bedeutung*, page 417–430. Radboud University Nijmegen.

Schmitt, Cristina; Holtheuer, Carolina; Miller, Karen (2004). Acquisition of copulas ser and estar in spanish: learning lexico-semantics, syntax and discourse. In *Proceedings of Boston University Conference on Language Development*, Somerville, MA. Cascadilla Press.

Schwenter, Scott A. (1994). The grammaticalization of an anterior in progress: Evidence from a peninsular spanish dialect. *Studies in Language*, **18**(1), 71–111.

Schwenter, Scott A. and Cacoullos, Rena Torres (2008). Defaults and indeterminacy in temporal grammaticalization: The 'perfect' road to perfective. *Language Variation and Change*, **20**, 1–39.

Schütze, Hinrich (1998). Automatic word sense discrimination. *Computational Linguistics. Special issue on word sense disambiguation*, **24**, 97–123.

Sebastián-Gallés, Núria (2000). *Lexesp: Lexico Informatizado Del Español*. Edicions Universitat Barcelona.

Seifert, Eva (1930). Haber y tener como expresiones de la posesión en español. *Revista de Filología Española*, **XVII**, 233–276;315–389.

Shibatani, M. (1991). Grammaticalization of topic into subject. In E. Traugott and B. Heine (eds.), *Approaches to Grammaticalization*, volume 2, pages 93–134. John Benjamins.

Singleton, Jenny and Newport, Elissa L. (2004). When learners surpass their models: the acquisition of american sign language from inconsistent input. *Cognitive Psychology*, **49**(4), 370–407.

Smith, Kenny and Wonnacott, Elizabeth (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, **116**, 444–449.

Sorace, Antonella (2000). Gradients in auxiliary selection with intransitive verbs. *Language*, **76**(4), 859–890.

Sorace, Antonella and Keller, Frank (2005). Gradience in linguistic data. *Lingua*, **115**, 1497–1524.

Swales, John (1990). *Genre Analysis*. Cambridge University Press, Cambridge.

Sánchez-Marco, Cristina and Marín, Rafael (2011). Generalising paths into psychological verbs. In *Going Romance XXV*.

Sánchez-Marco, Cristina; Boleda, Gemma; Fontana, Josep Maria; Domingo, Judith (2010). Annotation and Representation of a Diachronic Corpus of Spanish. In *Proceedings of Language Resources and Evaluation (LREC)*, Malta.

Sánchez-Marco, Cristina; Boleda, Gemma; Padró, Lluís (2011). Extending the tool, or how to annotate historical language varieties. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH)*, pages 1–9, Portland, OR, USA. Association for Computational Linguistics.

Sánchez-Marco, Cristina; Fontana, Josep Maria; Boleda, Gemma (to appear). Propuesta de codificación de la información paleográfica y lingüística para textos diacrónicos del español. uso del estándar tei. In *Nuevas perspectivas para la edición y el estudio de documentos hispánicos antiguos*. Peter Lang. Paper presented at CHARTA, Madrid 2009.

Sánchez Prieto, Pedro (2005). La normalización del castellano escrito en el siglo xiii. los caracteres de la lengua: grafías y fonemas. In *Historia de la lengua española*, page 199–213. Ariel, Barcelona.

Taylor, Ann (1994). The change from sov to svo in ancient greek. *Language Variation and Change*, **6**, 1–37.

Tenny, Carol (1994). *Aspectual Roles and the Syntax-Semantic Interface*. Kluwer Academic Publishers, Dordrecht.

Thielmann, Phillip (1885). Habere mit dem part. perf. pass. *Archiv für lateinische Lexikographie und Grammatik*, **2**, 372–423, 509–549.

Timberlake, Alan (1977). Reanalysis and actualization in syntactic change. In C. L. Li (ed.), *Mechanisms of syntactic change*, pages 141–177. University of Texas Press.

Torrens, M. Jesús (ed.) (2002). *Edición y estudio lingüístico del Fuero de Alcalá (Fuero viejo)*. Fundación Colegio del Rey, Alcalá de Henares.

Traugott, Elizabeth (1972). *The history of English syntax*. Holt, Rinehart and Winston, New York.

van Gelderen, Elly (2004). *Grammaticalization as economy*. John Benjamins.

Vendler, Zeno (1957). Verbs and times. *Philosophical Review*, **56**, 143–160.

Vlach, Frank (1993). Temporal adverbials, tenses, and the perfect. *Linguistics and Philosophy*, **19**, 231–283.

von Fintel, Kai (1995). The formal semantics of grammaticalization. In *North East Linguistics Society (NELS)*, volume 25, page 175–189.

Waltman, Franklin (ed.) (1999). *Textos y concordancias electronicos del Fuero general de Navarra*. Hispanic Seminary of Medieval Studies, Madison, WI.

Wang, William S. Y. (1969). Competing sound changes as a cause of residue. *Language*, **49**, 9–25.

Wang, William S. Y. and Chin-Chuang, Cheng (1977). Implementation of phonological change: The shuang-feng chinese case. In W. S. Y. Wang (ed.), *The lexicon in phonological change*, pages 148–158. Mouton, The Hague.

Webber, Bonnie (2009). Genre distinctions for discourse in the penn treebank. In *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP*, page 674–682, Suntec, Singapore.

Xu, Fei (2007). Rational statistical inference and cognitive development. In *The innate mind: Foundations and the future*, volume 3, pages 199–215. Oxford University Press.

Xu, Fei (2011). Rational constructivism, statistical inference, and core cognition. *Behavioral and Brain Sciences*, **34**, 151–152.

Xu, Fei and Griffiths, Thomas L. (2011). Probabilistic models of cognitive development. *Cognition*, **120**, 299–301.

Yllera, Alicia (1980). *Sintaxis histórica del verbo español: las perífrasis medievales*. Universidad de Zaragoza.

Yllera, Alicia (1999). Las perífrasis verbales de gerundio y participio. In I. Bosque and V. Demonte (eds.), *Gramática descriptiva de la lengua española*, chapter 52. Espasa-Calpe.

Zipf, George K. (1935). *The Psychobiology of Language*. Houghton-Mifflin.

Zipf, George K. (1949). *Human Behavior and the Principle of Least Effort*. Addison-Wesley.