2-9-2010

Formulaic Language and Adjective Categories in Eight Centuries of the Spanish Expression of 'Becoming' /quedar(se)/ + ADJ

Damián Wilson

Damián Chase Vergara Wilson
Candidate

Spanish and Portuguese
Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

, Chairperson

, Reader

FORMULAIC LANGUAGE AND ADJECTIVE CATEGORIES IN EIGHT CENTURIES OF THE SPANISH EXPRESSION OF 'BECOMING' /QUEDAR(SE)/ + ADJ

BY

DAMIÁN CHASE VERGARA WILSON

B.A., University Studies, University of New Mexico, 1999
M.A., Spanish, University of New Mexico, 2003

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Spanish and Portuguese

The University of New Mexico
Albuquerque, New Mexico

December 2009

Dedication

In memory of my grandfather Chase Altman Wilson who continuously gave me unwavering support in everything I have ever done. I am thankful that he was around long enough to share in the mutual delight of seeing this project's completion shine on the near horizon. I finished it for you Chase!

Acknowledgments

I was extremely fortunate to have had such an amazing dissertation committee and each member contributed significantly to this project. I thank Joan Bybee for initially having faith in my ability to carry out this type of research and for providing inspiration through her own investigations. I thank the Committee Chair, Catherine Travis for the hundreds of diligent written comments that encouraged the good aspects of the work while pointing out areas that needed improvement. I thank Clancy Clements for encouraging me to look deeper into the multiple earlier usages of the verb *quedar* and how it verb was intertwined with a similar one, *fincar*. I thank Anthony Cárdenas for all of his help in deciphering the earlier tokens and for being willing to sit on a linguistics committee.

I thank my wife, Jacqueline Flores, for her support. She inspired me to excel when things were good and reminded me that all I could do was take one step at a time when things were overwhelming. I thank my daughters for patiently waiting for the 'fun dad' to eventually emerge. I also thank my parents for their continual support and for being 'fun grandparents' at times when the 'fun dad' was stowed away.

No acknowledgments would be complete without thanking the awesome administrative staff at the Department. Martha Hurd, Kate Merril, and Rosario Johnson have always made sure that my loose ends were tied up.

Finally, I thank Enrique Lamadrid for being an eternal mentor and for helping me to get the Mellon Grant that made it possible for me to focus exclusively on this project for an entire year.

FORMULAIC LANGUAGE AND ADJECTIVE CATEGORIES IN EIGHT CENTURIES OF THE SPANISH EXPRESSION OF 'BECOMING' /QUEDAR(SE)/ + ADJ

BY

DAMIÁN CHASE VERGARA WILSON

ABSTRACT OF DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Spanish and Portuguese

The University of New Mexico
Albuquerque, New Mexico

December 2009

Formulaic language and adjective categories in eight centuries of the Spanish expression of 'becoming' *quedar(se)* + ADJ

by

Damián Chase Vergara Wilson

BA, University Studies, University of New Mexico, 1999
MA, Spanish Linguistics, University of New Mexico, 2005
Ph.D., Spanish Linguistics, University of New Mexico, 2009

Abstract

The purpose of this study is to track the diachronic development of exemplar clusters formed by the adjectives in the Spanish expression of becoming *quedar(se)* + ADJ (e.g. *quedar(se) solo* 'to be left alone', *quedar(se) espantado* 'to get scared'). This approach applies the same system of analysis used by Bybee & Eddington (2006) in their synchronic study of four verb + adjective combinations used to denote a change of state (*ponerse* + ADJ*, hacerse* + ADJ*, quedarse* + ADJ*,* and *volverse* + ADJ). Although there are diachronic studies that follow the development of constructions employing usage-based approaches (Israel 1996), and there are synchronic studies that apply the exemplar model to the analysis of constructions in Spanish (Bybee & Eddington 1996) this thesis is the first in-depth diachronic study known to the author that uses the exemplar model to account for the development of a construction over several centuries. It represents the continuation of preliminary studies by the author (Wilson 2006, 2009) but is much more in-depth, analyzing data from the 13th to the 19th centuries. A total of 1,374 tokens were analyzed in the time span indicated and were extracted from a corpus of 5,394,606 words

compiled from 119 narrative (or narrative-like) written works. This study provides evidence that, (a) frequently occurring verb + adjective combinations, such as prefabs, serve as the central members of exemplar categories, (b) formulaic combinations, such as prefabs, have longevity, (c) the categories to which formulaic sequences belong have longevity, (d) categories mutate over time by becoming more centralized, changing central members, expanding to new types or by contracting as types are lost, and (e) there is a correlation between the token frequency of the central member of a category and the type frequency of that category. The data studied in this thesis suggest that the rise in overall standardized frequency of *quedar(se)* + ADJ has to do with the decrease in frequency of a similar expression of 'becoming', *fincar(se)* + ADJ, as many adjectives associated with the latter come to be used with the former. In tracking the continuum of adjective categories over time, this investigation provides insight into diachronic trends of formulaic language showing trends of emergence as categories expand, contraction as forms fall into disuse, and stability as some categories change very little over time.

Table of contents

List of figures

List of tables

List of charts

Chapter 1: Introduction

In Spanish, there are several verb + adjective combinations that function as constructions used to indicate a change of state. These constructions have been studied by various investigators (Crespo 1949, Coste & Redondo 1965, Fente 1970, and Eberenz 1985, Eddington 1999, 2002, Bybee & Eddington 2006) and have received commentary in reference grammars (Butt & Benjamin 2004) as attempts have been made to illustrate how a particular construction is used and to explain the patterns of adjective used with each verb (this is discussed in more detail in Section 1.2).

When applied to written and spoken date, Bybee & Eddington (2006) found these attempts at explaining the distribution of adjectives among verbs of 'becoming' do not account for usage. The present study uses Bybee & Eddington (2006) as a point of departure. It is a usage-based study of the distribution of adjectives among certain verbs of 'becoming'. Using the exemplar model to account for the distribution of adjectives in expressions of 'becoming', Bybee & Eddington (2006) studied the four most frequent verb + adjective constructions used in Modern Spanish change-of-state expressions with animate subjects: *quedarse* + ADJ*, ponerse* + ADJ*, hacerse* + ADJ, and *volverse* + ADJ. These four verbs occur frequently in Spanish and, when not used as 'become' verbs, are used most typically to mean the following: *quedarse* 'to remain, to stay'*, poner* 'to put'*, hacer* 'to make' and *volver* 'to return, to turn'. Despite their different meanings when used otherwise, used with adjectives in change-of-state constructions, these verbs mean 'to become'. There are factors that make these verb + adjective combinations ideal to study through the framework of a usage-based model. Instead of discrete rules that are not only hard to apply, but also fail to account for the distribution adjectives in these

constructions, Bybee & Eddington found that the best way to account for the distribution of certain adjectives with certain verbs was through a usage-based account that considers speaker experience. Among Bybee & Eddington's (2006: 323) main findings were that the categories of adjectives used in frequently occurring verbs of 'becoming' in their study (*quedarse* and *ponerse*) demonstrated "dense clusters of semantically related adjectives centered on a high-frequency exemplar", and that "novel instances of verb + adjective sequences are based on analogies to previous experience and not on rules that refer to abstract features". Their study is described in more detail in Section 1.2.2 of this chapter and is referenced frequently throughout this dissertation.

This investigation takes the stance that the high degree of organization that Bybee & Eddington (2006) observed must be the result of an ongoing process in which previous usage affects subsequent usage. Preliminary investigations on *quedar(se)* + ADJ (Wilson 2006, 2007, 2009) have shown that some of the adjective categories proposed in Bybee & Eddington (2006) have either existed as categories since at least the 1200's, or have emerged at some point in time since then. Since the aforementioned preliminary investigations have shown that adjective categories stemming from *quedar(se)* + ADJ are based on analogies to previous experience, this study seeks to reveal in detail what these previous linguistic experiences were and how they have changed over the centuries. Moreover, in some cases this study seeks to show continuity of some of these linguistic experiences by demonstrating that, instead of mutating wildly over time, some of these clusters of semantically related adjectives have been centered on a high frequency exemplar for many centuries. In other words, this study seeks to reveal sites where there

is relatively little change over the centuries as well as to reveal the emergence and development of new patterns of adjective distribution.

The purpose of this chapter is to introduce relevant issues to the construction *quedar(se)* + ADJ that will serve as a background for the ensuing diachronic analysis. Section 1.1 touches upon the various ways in which a change of state can be expressed in Spanish. This section also introduces basic concepts of the exemplar model and construction grammar as adequate approaches for analyzing the present data (these are presented more thoroughly in Chapter 2). Section 1.2 provides the reader with a review of previous investigations of change-of-state constructions, including a discussion of some of the works that proposed rigidly discrete criteria in order to account for the usage of certain adjectives in combination with certain verbs (Coste & Redondo 1965, Crespo 1949, Eberenz 1985, Fente 1970). This section also discusses Eddington's previous investigations (1999, 2002), gives a more in-depth presentation of Bybee & Eddington's (2006) findings, and outlines the results of my own previous studies. The last section, Section 1.3, presents some of the findings and conclusions of the present study and gives the reader an overview of the rest of the document.

1.1 'Becoming' in Spanish

In order to understand why the construction *quedar(se)* + ADJ presents an ideal phenomenon for a diachronic, usage-based investigation, it is necessary to compare this change of state construction to other ways of expressing a change of state in Spanish. There is a variety of ways to expresses 'becoming' in Spanish through both multi-word constructions such as *quedar(se)* + ADJ or *poner(se)* + ADJ, and through any of the numerous verbs that inherently express a change of state. Commenting upon this variety

of ways to express 'becoming', Eddington (1999, 2002) wrote that there has not been a satisfactory treatment of change-of-state expressions in any reference grammars on Spanish and cites Fente (1970: 157-8) as having searched a plethora of widely used reference grammars on this topic, coming up with no results. Reference grammars such as Butt & Benjamin (2004), Eddington observes (1999, 2002), have chiefly relied on the information from Fente's article, and from three others that discuss expressions of 'becoming' (Coste & Redondo 1965, Crespo 1949, Eberenz 1985). Part of the problem in explaining verbs of 'becoming' in Spanish is attributable to the variety of means by which one can express a change of state: verbs that mean 'to become' (e.g. *transformarse, convertirse*), middle-voice verbs that indicate a change of state (e.g. *enfermarse* 'to get sick', *emborracharse* 'to get drunk'), and constructions that have emerged in Spanish that mean 'to become' in the context of a change-of-state construction (e.g. *quedarse* + ADJ, *ponerse* + ADJ, etc).

There are verbs in Spanish that literally mean 'to become', the two most common of which are *convertirse* and *transformarse* in Modern Spanish. These two verbs are used almost exclusively with nominal complements (Eddington 1999 & 2002). The extensive online dictionary *WordReference.com* (Kellog 1999) gives the following suggested translations for *convertirse*: 'to turn into, become', and the following for *transformarse*: 'to change, to turn into'. These verbs can be used both transitively (1), and intransitively with *se* (2) and (3).

(1) *Convertí / transformé el salón en una discoteca.*
'I turned the living room into a discotheque.'

(2) *Se convirtió en un idiota.*
'He / she turned into an idiot.'

(3) *Los harapos se transformaron en un vestido espléndido*
'The rags transformed into a splendid dress.'

The expression *llegar a ser*, and the verbs *hacerse* and *volverse,* are used to indicate a change of state and can be used with both nominal (exs. (4), (5), and (6)) and adjectival complements (exs. (7), (8), and (9)) (Eddington 1999 & 2002).

(4) *Llegaron a ser dueños de un rancho.*
'They became owners of a ranch.'

(5) *El muchacho ya se volvía hombre.*
'The boy was turning into a man.'

(6) *Ramón se hizo soldado.*
'Ramón turned into a soldier.'

(7) *Llegaron a ser famosos en la aldea.*
'They became famous in the village.'

(8) *El muchacho ya se volvía loco.*
'The boy was going crazy.'

(9) *Ramón se hizo rico.*
'Ramón got rich.'

Another common way to express a change of state is with middle-voice verbs that describe a specific verbal situation in which the subject went from one state into another. Eddington (1999: 24), although excluding them from his data set, refers to these verbs as "verbs that express both the change and result of the change". Fente (1970: 169) and Crespo (1949: 212) in their outlines of verbs of 'becoming' in Spanish recognized these verbs, mentioning that there is a set of 'reflexive' verbs that inherently express a change of state. However, the term 'reflexive' can be problematic. Many intransitive verbs are generically called reflexive because they appear with a pronoun. Kemmer (1993, 2003) points out that, despite similarities in form of the reflexive and middle voice as seen in Spanish (both use the pronoun *se*), there is a distinction between reflexive usages of verbs

and the middle voice, which expresses a change of state. The difference between these two usages is described according to the degree in which participants involved in the verbal situation are conceptually distinct from one another, a parameter known as the "relative distinguishability of participants" (Kemmer 2003: 108). The reflexive has two different participant roles that are carried out by the same entity, as in (10). In the middle voice (Kemmer 1993, 2003), there is no such distinction of roles and there is only one participant who initiates the action, and is affected by it, as in (11).

(10) *Se cortó.*
'He / she cut him / herself.'

(11) *Se enfadó.*
'He / she got mad.'

Other examples of the middle voice are found in such verbs as: *enfermarse*, 'to get sick', *emborracharse*, 'to get drunk', *inspirarse*, 'to get inspired', and *perderse*, 'to get lost'. The middle voice is very widespread in Spanish and adds to the repertoire of expressions of 'becoming' available to speakers.

However, this does not eliminate other possibilities in Spanish for describing a change of state. Specifically, many of the changes of states described by the middle voice could also be described in verb + adjective combinations involving the four verbs of 'becoming' studied by Eddington (1999, 2002), and later by Bybee & Eddington (2006): *quedarse, hacerse, ponerse,* and *volverse*. Relying on an analysis of invented examples, Eberenz (1985: 464) mentions that a speaker, due to stylistic questions, could alternate between *hacerse viejo* and *envejecerse*, 'to get old', *volverse pobre* and *empobrecerse*, 'to become poor', and between *ponerse animado* and *animarse*, 'to get lively' with no significant change in meaning. Analyzing contexts of usage extracted from novels, Fente

(1970: 169-70) offers that the expressions *se cansó*, 'he/she got tired', *se perdió*, 'he/she got lost', and *se rompió*, 'it broke', could synonymously be expressed with verb + past participles combinations involving *quedarse* + *cansado* 'tired', *perdido* 'lost', or *roto* 'broken'. Fente not only claims that these particular past participles could not be employed with *hacerse, ponerse,* and *volverse*, but does not acknowledge any overlap in which an adjective occurs with multiple change-of-state verbs. Responding to Fente's claim, Eddington's work (1999, 2002, Bybee & Eddington 2006) documented that there was overlap in usage. This is relevant to the present study because the model for this study, Bybee & Eddington (2006), and previous diachronic studies by the present author (Wilson 2006, 2007, 2009) have shown that there are no discrete categories that feed rules of adjective usage in change-of-state construction. Instead, categories of adjectives used in expressions of 'becoming' are based on flexible categories formed diachronically through previous usage, which might lead to an overlap of adjective use.

The rest of this section provides a brief introduction to the exemplar model and to usage-based construction grammar, which form the theoretical underpinnings for this dissertation and for the study that inspired it (Bybee & Eddington 2006). These models have proven to be perfectly adaptable to historical data as they allow for linguistic change in a diachronic setting, as well as for the endurance of conventionalized forms. Chapter 2 will describe these two models in more detail and discuss issues of their application to diachronic linguistic research

The exemplar model is a usage based model of cognitive representation in which categorization is proposed to be an ongoing process that comes about through specific instances, episodes, and experiences (Bybee 2001, 2006). This applies as much to other

cognitive processes as it does to language, in which a person's experience with language plays a central role in the formation of categories of linguistic material. Speakers are constantly involved in the production and reception of a variety of modes of language. All of this linguistic material is mapped onto matching representations in the speaker's mind, strengthening those representations (Bybee & Eddington 2006, Chandler 2002, Pierrehumbert 2001). Episodic information is also stored along with many abstract levels of linguistic information. These collective representations are called exemplars and the individual occurrences of a linguistic item are considered tokens. Categories of exemplars show gradient membership in which highly frequent exemplars are the central members of categories and less frequent ones are more marginal. If the speaker encounters a token that does not map onto an existing category of exemplars, it is evaluated on the basis of similarity to existing exemplars and represented as being close to, or far from, existing categories. This judgment of similarity and subsequent categorization can be based on metaphorical or analogical similarities to established categories. As it is a model based on speaker experience, categories are constantly being updated to reflect ongoing experiences with a certain item.

Since present usage is based on the accumulation of previous experience, one of the goals of this study is to portray the way that exemplar clusters are formed over a long period of time. Viewed this way, this research seeks to portray the construction *quedar(se)* + ADJ as part of an ever-evolving continuum. As Bybee & Eddington (2006: 327) stated, the model they have chosen "allows representations to change as language is used because every token of experience has an impact on cognitive representation". I

seek to show how these representations have changed and how they have remained the same over time.

In a similar way to the exemplar model, usage-based construction grammar operates from the standpoint that grammatical forms emerge from usage. One of the basic notions of construction grammar is that constructions are considered to be the basic, elemental units of language (Croft 2001, Goldberg 1995, 2003). Goldberg (2003: 219) defines constructions as "stored pairings of form and function, including morphemes, words, idioms, partially lexically filled and fully general linguistic patterns."

Constructions have fixed units that characterize the construction and may also have an open slot that is filled by a certain class of lexical item (Bybee & Eddington 2006, Erman & Warren 2000). In the case of the construction, *quedar(se)* + ADJ, the verb *quedar(se)*, in all of its morphological realizations, would constitute the fixed unit while the open slot is represented by the notation ADJ and is filled by adjectives and prepositional phrases. This study, following Bybee & Eddington (2006), will focus on the categories formed by the adjectives that fill the open slot of the construction at hand.

Furthermore, it is important to study constructions from the perspective that, taken as a whole, they have a meaning that goes beyond the lexical meaning of the individual parts; constructions are direct form-meaning pairs (Goldberg 1995 & 2003, Lakoff 1987). In the case of *quedarse* + ADJ, proponents of construction grammar would recognize that the construction is analyzed through context by speakers as meaning 'to become' and the adjective determines the state resulting from the change. The construction *quedar(se)* + ADJ must, therefore, be analyzed as a whole, and independently of what the individual parts would mean outside of the construction.

The construction that this study focuses on is the verb + adjective combination *quedar(se)* + ADJ. The parentheses indicate the variable usage of a 'reflexive' pronoun in the construction (1s, *me*; 2s, *te*; 2s/3s, *se*; 1p, *nos*; 2p *os;* 2p/3p, *se*), which may (*me quedare* (12)) or may not (*quedara* (13)) be used. This doesn't appear to be the case with the other verb + adjective combinations listed by Bybee & Eddington (2006), *ponerse* + ADJ*, hacerse* + ADJ, and *volverse* + ADJ, which categorically occur with the reflexive pronoun when used as change-of-state constructions. However, when commenting upon the findings of other researchers, such as Bybee & Eddington (2006), this study will omit the parentheses where it more accurately portrays the other researchers' manner of notating the constructions described.

When the capitalized abbreviation ADJ is used with a Spanish verbal infinitive (e.g. *quedar(se)* + ADJ), it is an indication that the combination is a change-of-state construction. Furthermore, the adjective (ADJ) represents an open slot in which many adjectives can be used, whereas the verb is a fixed element. The frequently used term 'verb + adjective combination' is used to denote these combinations in a general sense without necessarily commenting upon their status as a change-of-state construction.

(12) ***Sol****. He menester camisas, jubon, casaca, calçones, médias, çapatos, y sombrero, y con quatro pesos no áy para todo ... quiero jugar, quiçà ganarè para comprarlo todo nuevo.*
Sar. *Y si el dado dize mal?*
Sol. *Me quedare sin lo uno, y sin lo otro, y entonces dire; desnudo nacì, desnudo me hallo, y desnudo morirè.*
'**Soldier**. I need shirts, soap, jacket, pants, socks, shoes, and a hat, and with four *pesos* there isn't enough for all … I want to play, maybe I'll earn enough to buy everything new.
Sergeant. And if the dice comes out badly?
Soldier. I shall be left without the one, or the other, and then I will say; naked I was born, naked I find myself, and naked I shall die.' (*Diálogos nuevos en español y en francés*. 1708. Francisco Sobrino. Lemir)

(13) *Y que si no muriera quedara con el cuello tuerto en postura de ahorcado.*

'And if he didn't die he would be left with his neck twisted into the posture of a hanged man.' (*Comedia Vidriana*. 1525?. Jaime de Huete. Lemir)

1.2 Previous attempts at explaining change-of-state construction in Spanish

One of the purposes of Fente's (1970) article was to propose guidelines for teaching students about the rules used to link a verb used in an expression of 'becoming' to the appropriate adjective and / or situation. A few years later Eberenz (1985: 463) lamented that even though verbs of 'becoming' present a potentially huge area of research, "change of state verbs are … the Cinderella of the functional verbs of Spanish". This section focuses on the criteria proposed by the small body of research that has sought to explain the distribution of adjectives in verb + adjective combinations used in Spanish to express a change of state (Crespo 1949, Coste & Redondo 1965, Eddington 1999, 2002, Eberenz 1985, Fente 1970). Furthermore, this section touches upon the article that serves as a point of departure for the present study- Bybee & Eddington (2006).

1.2.1 Two studies by Eddington (1999 & 2002)

Given that there is a plethora of verb + adjective combinations in Spanish that are used to express the relatively synonymous idea of a change of state, the question has arisen as to what factors influence a speaker to choose one combination over another. Four previous studies on expressions of 'becoming' (Crespo 1949, Coste & Redondo 1965, Fente 1970, Eberenz 1985) presented a major disadvantage in that they didn't apply all of the criteria consistently to all the verbs studied, as observed by Eddington (1999). Also, these studies didn't all study the same verbs. Crespo (1949), for example,

considers *ponerse*, *hacerse*, *volverse*, and *llegar a ser*, but not *quedarse* in his list of verbs of 'becoming' in Spanish which was found by Bybee & Eddington (2006) to be the most frequent verb used with an adjective to express a change of state.

These previous studies posited, in some cases, discrete criteria for choosing among various combinations. For example, according to Crespo (1949: 210), *ponerse* is "the only one of the four verbs that is used to express transitory changes in the physical, emotional, or mental state of persons or animals" (e.g. *ponerse triste / pálido* 'to become sad / pale'). In other cases, the authors were unable to formulate a clear proposal as to the criteria for usage regarding some of the verbs. Fente (1970: 162) writes that "it is difficult to establish what types of adjectives may accompany the verb *hacerse*" giving only the ideas of 'overcoming' something or comparison as the criteria that would necessitate the usage of this verb (e.g. *hacerse rico* 'to become rich').

Eddington (1999) used the four previous studies of expressions of 'becoming' (Crespo 1949, Coste & Redondo 1965, Fente 1970, Eberenz 1985) in order to develop viable criteria for selecting verb + adjective / noun combinations that could be tested in data from spoken and written corpora. He applied four of the most concrete syntactic / semantic factors from these previous works to 1,283 tokens of the following change of state verbs: *llegar a ser*, *ponerse*, *volverse*, *quedarse*, *convertirse*, *transformarse*, and *hacerse*. The factors applied were 1) whether the verb's complement is a noun or an adjective, 2) whether the change is incremental or sudden, 3) whether the change occurs actively or passively, and 4) whether the noun or adjectival complement typically occurs with the copular *ser* or *estar* (both mean 'to be' but *estar* is assumed describe temporary states while *ser* is deemed to describe more permanent ones). He found that these

semantic and syntactic features did not effectively determine the choice of a certain verb + adjective combination in a given situation. His analysis demonstrated that there was a great deal of overlap of adjectival complements between the verbs and that the proposed criteria could not successfully account for the distribution of adjectives among the seven verbs studied.

Nevertheless, a key finding of Eddington (1999) is that certain combinations were lexically determined in the sense that a certain adjective would always call for a certain verb. In his corpus, consisting of written and spoken data, he found that there were noticeable tendencies in which some adjectives appeared exclusively with one verb. For example, the adjective *dormido* 'asleep' appeared 28 times in his data always with the verb *quedarse*. Eddington also found that while there was a "great deal of overlap and encroachment on the uses of each verb" (1999:33), at the same time, there was a great deal of consistency. Among 163 tokens of adjectives that appeared more than once in the data, 76 always occurred with the same verb while 87 occurred with different verbs. In light of this, Eddington proposed that many verb + adjective combinations could be lexically determined. Similarly, but in an opposite vein, he proposed that combinations that involved adjectives that appeared with more than one verb were not lexically determined. Their usage would have to be influenced by some other factor. In an attempt to explain a possible influence, Eddington (1999: 33) went on to propose that the overlap may have been due to dialectal differences as his data came from both a spoken corpus based on Peninsular Spanish, and a written corpus which included authors from several different countries.

In order to develop his research on change of state verbs, Eddington (2002) presented a questionnaire to native Spanish speakers with the task of deciding which change-of-state expression speakers perceived to best finish a given sentence, as in Example (14). These expressions contained a constant adjective that was combined with each of the seven change of state verbs mentioned above. The material in the questionnaire was designed to test the same four semantic / syntactic factors with the same verbs as the previous analysis (Eddington 1999). Examining (14), for example, the complement *urgente* 'urgent' is adjectival and would be used in another context with the copular verb *ser*. It also represents a quick change that was involuntary. Also, this questionnaire could test for overlap. If all speakers chose *quedar(se)* for a specific adjective, it would be evidence against overlap. If they chose different verbs, it is evidence that a certain adjective overlaps.

(14) *Por la cara que ponía y su tono de voz, nos dimos cuenta de que ya no podíamos más. La situación...*
("Judging by his appearance and tone of voice, we realized that we couldn't wait any more. The situation...)
__*se había hecho urgente.* ("had become urgent")
__*se había convertido en urgente.*
__*se había transformado en urgente.*
__*se había puesto urgente.*
__*se había vuelto en urgente.*
__*había llegado a ser urgente.*
__*se había quedado urgente*.
(Eddington 2002: 922, Ex 1. Format of original example replicated here)

The results of the newer study (Eddington 2002) supported Eddington's earlier finding that even though there is overlap in the usage of verbs and adjectives with these expressions of 'becoming', there is also great regularity. Because the participants were almost all from the region of Andalucía, Spain, this refutes the proposal in his previous study that overlap was due to dialectal differences. Despite the regularity of verb +

adjective combinations, he concluded that the criteria originally tested in Eddington 1999 were not applied by speakers as they chose which verb to use in a given context.

1.2.2 Overview of Bybee & Eddington 2006

Continuing the previous investigations of verbs of 'becoming' by Eddington (1999, 2002), Bybee & Eddington (2006) studied the categorization of adjectives used in combination with the following four verbs and an animate subject: *ponerse*, *volverse*, *quedarse*, and *hacerse.* They extracted 423 tokens from a spoken corpus from Spain and written data from fifteen novels from various Spanish speaking countries; the total of the written and spoken corpora was approximately 2 million words. The goal of their article was to develop and apply a model that would adequately account for the categorization of verb + adjective combinations in Spanish based on linguistic usage and speaker experience. This study also represents a shift in the orientation of studying verbs of 'becoming' because, instead of trying to reveal inherent characteristics of the individual verbs or adjectives that would account for usage, Bybee & Eddington (2006) focused on the distribution of adjectives among the 'becoming' verbs and the way in which speakers categorized these adjectives into semantically related clusters.

They noted that even though there are strong tendencies observable in the data where certain adjectives always occurred with the same verb, "the boundaries of the categories of the complements to these verbs are fuzzy" (Bybee & Eddington 2006: 324). Instead, the distribution of adjectives in these constructions displays a graded category membership where frequently occurring combinations are more central and less frequent ones are more marginal. With this in mind, they determined that it would be best to study the usage of verbal complements with a cognitive model of representation that would

adequately explain this graded category membership. They chose a usage-based exemplar approach because it is based on the speaker's experience with language and can account for strong tendencies of verb + adjective combinations while allowing for the observed overlap of adjective use with different verbs.

Among their main findings was the following: "novel instances of verb + adjective sequences are based on analogies to previous experience and not on rules that refer to abstract features" (Bybee & Eddington 2006: 323. Instead, the results of their research support an exemplar model of representation where categories are formed based on the speaker's previous experience of language use. Frequent exemplars are central category members because they have stronger representation, yet frequent and marginal members may serve as the basis for analogical extension into novel linguistic usage.

In order to create a speaker-based organizational structure of adjectival exemplars into exemplar clusters, a graduate student and native speaker of Mexican Spanish analyzed cards with the different adjectives found in the data written on them. She then arranged the cards into groups according to perceived semantic similarity; the closer the cards, the closer the meaning. The validity of the proposed organizational structure of exemplar clusters was supported by two later experiments conducted with numerous native speakers in Spain: a similarity experiment and an acceptability experiment. These are presented respectively in Sections 1.2.2.1and 1.2.2.2 below.

Observing that some verb + adjectives combinations appear in the data with relatively high token frequency, they proposed that these high-frequency combinations are prefabs as in Examples (15), (16), & (17) below from their study.

(15) *quedarse solo* (n=28)
'to be left alone / to end up alone'

(16) *quedarse inmóvil* (n=17)
'to be immobile'

(17) *ponerse nervioso* (n=17)
'to get nervous'

The organizational structure of the exemplar clusters shows prototype effects in which high-frequency prefabs serve as central members of the clusters where other members are more marginal. As will be confirmed in this diachronic study, central members serve as a model upon which newer, more marginal members are added to a category over time. For example, in talking about this research Bybee (2006: 727) gives the example of the prefab *quedarse inmóvil*, commenting that it "seems to have spawned many similar uses with other adjectives". In Bybee & Eddington (2006) 15 other adjectives appeared in the clusters centering on *inmóvil* that had a similar meaning to this central member; these occurred only one to three times each in their data and were, therefore, considered marginal members. These more marginal members share semantic characteristics with the central member on a variety of levels. Some are actual synonyms to the central members, *inmóvil*, as in the case of *parado* 'stopped, standing'. Figurative or metaphorical expressions are included, such as *de piedra* 'of stone'. Also, some members appear in this category that share some features with *inmóvil* but have additional features as in the adjective *atrapado* 'trapped' (Bybee 2006). Whether or not the peripheral members are synonyms to one another, they are linked semantically in some way to the central member.

Bybee & Eddington present three types of evidence that support a model of categorization in which the exemplar with the highest frequency is the central member around which other members are organized. First, the most productive categories with the adjectives that had the highest token frequency displayed the type of organization

discussed above in which marginal members were clearly centered on a high-frequency adjective. They proposed that, in most cases, productivity of new adjectives was based on analogy to an already established frequent member. One of the things that this dissertation seeks to do is to illustrate this productivity diachronically as new usage emerges out of old. Of the four verbs studied in Bybee & Eddington (2006), only *quedarse* and *ponerse* consistently displayed a strong tendency of centralization whereby categories were clearly organized around a high frequency member. The other two verbs, *hacerse* and *volverse*, appeared with markedly lower frequency in the data and did not display the highly centralized category organization seen in the first two verbs; instead, they had a "much more scattered and miscellaneous distribution with adjectives" (Bybee 2006: 727). It would appear that the low frequency of these two verbs in the data could account for the lack of centralized category organization, or vice versa.

The second piece of evidence is that the categorization of adjectives showed a family resemblance in its organization in which the central member served as the category member that linked other members. Without the central member, these more marginal members may not necessarily be linked. Bybee (2006: 727) gives the example of *inmóvil* 'immobile' serving as the link between *petrificado* 'petrified' and *atrapado* 'trapped': *petrificado - inmóvil - atrapado*. Without *inmóvil*, she writes, it would be difficult to propose a semantic link between *petrificado* and *atrapado*.

The third type of evidence that supports the model of categorization in which adjectives are organized around a high frequency central member comes from the two experiments conducted on native speakers of Spanish in Spain: a similarity experiment and an acceptability experiment. These are outlined below.

1.2.2.1 The similarity experiment

In the similarity experiment, participants were given a questionnaire that included all possible pairings of 20 adjectives from the researchers' proposed exemplar clusters: seven from the *quedarse inmóvil* clusters, seven from the *quedarse sorprendido* 'to be surprised' clusters, and six from the *hacerse* + adjective clusters. This resulted in 190 test items that were divided into two questionnaires of 95 pairings apiece shown without surrounding context (e.g. *inmóvil, petrificado* 'motionless, petrified'). Seventy-seven participants were instructed to rate the similarity of each pair on a five point scale. These ratings were used in a multidimensional scaling (MDS) analysis that placed the adjectives in conceptual proximity to one another reflecting their reported similarities or differences. Figure 1 resulted from the MDS analysis in Bybee & Eddington (2006: 349).

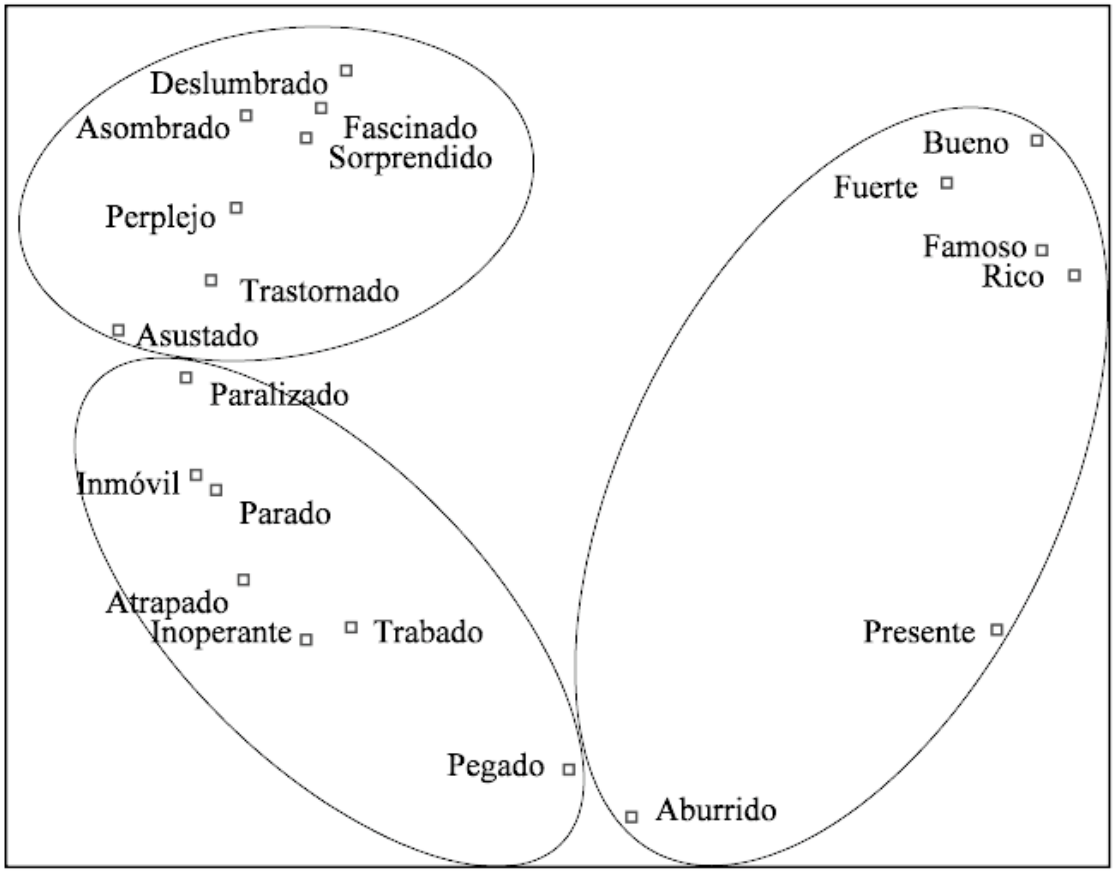


Figure 1: Multidimensional scaling of ratings from the similarity experiment; Bybee & Eddington (2006: 349)

The ovals in Figure 1 were drawn by Bybee & Eddington (2006) after the MDS configuration had been generated. These ovals represent the exemplar clusters formed by the researchers in the earlier part of their research, based on the organization of the adjectives originally proposed by the native Spanish speaker graduate student. It was somewhat surprising that *inmóvil* did not occupy a more central position as the researchers expected and that speakers had placed *famoso* 'famous', *rico* 'rich', *bueno* 'good', and *fuerte* 'strong' together, unlike the graduate student consultant had done. However, the multidimensional scaling analysis did put *sorprendido* in the center of

related adjectives and highly resembles the exemplar clusters previously formed by Bybee and Eddington's native speaker participant. Even though *inmóvil* does not occupy a central position, the adjectives proposed to appear in its exemplar clusters appeared nearby in a way that made it possible to enclose them in an oval without intruding on other exemplar clusters represented in Figure 1. Furthermore, despite the fact that the adjectives from the *hacerse* + adjective clusters weren't organized exactly as they were by the researchers, they were generally in proximity to one another and didn't intrude upon other clusters. In fact, based on the placement of the adjectives according to the MDS analysis, a vertical line would separate the *quedarse* adjectives (the ones appearing in the two ovals to the left) from the *hacerse* adjectives (appearing in the oval to the right). Taken as a whole, this experiment supports the groupings of adjectives into the proposed exemplar clusters[1].

1.2.2.2 The acceptability experiment

The acceptability experiment in Bybee and Eddington (2006) was designed to determine whether the frequency of exemplars and the semantic similarity proposed by the researchers would have correlations to acceptability ratings done by native speakers. Here the researchers sought to test the hypothesis that "stronger exemplars serve as the basis for the production of novel expressions" (Bybee & Eddington 2006: 349) with an experiment that could determine whether or not the experience that speakers have with certain verb + adjective combinations was part of their overall linguistic knowledge.

[1] An MDS study was conducted on adjectives found in the data for this thesis. The results supported the proposed category structures. See Chapter 6 for more details.

A group of 48 college students in Murcia, Spain, were given a questionnaire that had 31 examples consisting of *quedarse* and *ponerse* extracted from Bybee and Eddington's (2006) corpora of expressions of 'becoming'. Most of these were presented in the original context in which they occurred, although some had to be modified slightly to fit the format of the questionnaire. Also, the verb + adjective combinations were underlined in order focus attention on the expression of 'becoming' and minimize influence from the rest of the sentence on the participant's judgment. The questionnaire gave the participants the task of rating the stimuli on a five-point Lickert-type scale from *perfectamente bien* 'perfectly fine' to *raro* 'strange / odd / rare' as in (18) below (taken from Bybee & Eddington 2006: 350).

(18) *Con el vino y la voz pegajosa <u>se estaba poniendo sentimental</u>.*
'With wine and an over-affectionate voice he was getting sentimental'
Perfectamente bien Raro
\+ -------- + -------- + -------- + -------- +
'Perfectly fine' 'Odd'

The questionnaire presented three types of stimuli chosen on the following criteria: "(i) items that contained the high-frequency adjectives that make up the core of the exemplar clusters, (ii) items containing low-frequency adjectives that are semantically related to one of the core clusters of each verb, and (iii) low frequency adjectives that are semantically unrelated to other uses of the verb" (Bybee & Eddington 2006: 350) as exemplified respectively in (19), (20), and (21).

(19) High-frequency: *quedarse quietos* 'to be still', *ponerse nervioso* 'to get nervous'

(20) Low-frequency related: *quedarse paralizada* 'to become paralyzed' (related to central member *inmóvil* 'motionless'), *ponerse inaguantable* 'to get intolerable' (related to central member *pesado* 'annoying')

(21) Low-frequency unrelated: *quedarse orgullosísimo* 'to be proud', *ponerse viejo* 'to get old'

The researchers found that high frequency items were rated as most acceptable. Even though participants tended to rate high-frequency and low-frequency related stimuli as belonging on the perfectly fine side of the scale, there was a statistically significant difference between these two based on frequency. The most notable difference, however, was the one between low-frequency related items and low-frequency unrelated items indicating that the latter were much less acceptable than the former.

This experiment provides evidence that "acceptability is a function of experience" (Bybee & Eddington 2006: 352) and supports the patterns of exemplar categorization proposed earlier in the article. The fact that high-frequency items were judged to be the most acceptable supports the proposition that not only is frequency represented in the speaker's conception of language, but that it is also a factor in determining acceptability. The fact that low-frequency related items were judged as more acceptable than low-frequency unrelated items shows that semantic relatedness is also a key factor in acceptability judgments. Even if a certain verb + adjective combination has never occurred before, it will be more acceptable if it is semantically similar to a high frequency item. The finding that low-frequency related items were rated to be more acceptable than low-frequency unrelated items supports the centralized structure of exemplar categorization in which the central member is a high-frequency adjective surrounded by other semantically related ones.

1.2.3 Preliminary diachronic investigations of *quedar(se)* +ADJ

This section presents commentary on the relevance of previous studies that explored the construction *quedar(se)* + ADJ in a diachronic format using the exemplar

model (e.g. Wilson 2006, 2007, 2009). These studies represent an attempt to connect previous usage with present usage of adjectives used in *quedar(se)* + ADJ. They laid the groundwork for the present study by revealing many of the diachronic trends examined in detail in this study.

Examining data from the 13th, 15th, 17th, & 19th centuries, Wilson (2009) looked at changes in the exemplar categories formed by the open 'adjective' slot in two ongoing sets of exemplar clusters: the *quedar(se) solo,* 'alone' clusters, and the *quedar(se) confuso / sorprendido,* 'confused / surprised' clusters. As with Bybee & Eddington (2006), this study employed the exemplar model to account for the categorization of adjectives, and to account for the increased productivity of this change-of-state construction over time.

By investigating the categorization of adjectives in the construction *quedar(se)* + ADJ with a human subject, these preliminary studies have provided evidence that: (a) types with relatively high token frequency serve as the central members of exemplar categories, (b) central members of exemplar categories may retain their status over many centuries, (c) prefabs have longevity, and, (d) categories may change over time in a variety of ways. The changes in categories observed in the data show that categories can become more centralized around a member with high token frequency. This status of central member may also pass from one adjective to another during the course of time as the token frequency of the central member is surpassed by a related type.

At the same time, these preliminary results pointed to the need to execute a more in-depth study that the present thesis has sought to implement. In order to develop a more accurate portrayal of adjective use, this study examines data from every century from the

1200's to the 1800's, with commentary on the 1900's using data from Bybee & Eddington (2006) and Davies (2002-). This is described in the following section.

1.3 Summary of main points

By examining written and spoken data of change-of-state verbs, Bybee & Eddington (2006) found that the most plausible explanation of adjective distribution among these verbs lay in a model that recognized speaker experience and cognition as central factors in production and interpretation. Because of its flexibility, the exemplar model is ideal in accounting for adjective distribution. Through usage, speakers develop categories of similar items that show gradient structure in which frequently experienced exemplars are the central members. In the input, speakers decode the meaning of incoming items based on a comparison of incoming items to ones that have been experienced. In the production of language, speakers draw from these categories, either using previously experienced combinations or developing novel ones through analysis to existing ones.

The present study looks at one of the change-of-state expressions studied by Bybee & Eddington (2009), *quedar(se)* + ADJ, and examines the categories of adjectives found in this construction from the 1200's to the 1800's. One of the goals is to test whether this type of analysis can be applied to diachronic data. Bybee & Eddington performed a synchronic analysis of the data while recognizing the fact that previous usage affects subsequent usage. Furthermore, they proposed that their model of analysis could serve as a predictor of subsequent usage. In order to capture the ongoing trends of adjective usage in *quedar(se)* + ADJ through time, this study looks at every century in the time span studied, unlike Wilson (2009) which looked at alternating centuries. This

study shows that by the time Bybee & Eddington performed their analysis some categories were becoming more productive (e.g. adjectives related to *quedar(se) solo* 'to be left alone'), with the central member gaining in token frequency and the category, as a whole, extending to new types. Other categories were in the process of becoming less productive (e.g. adjectives related to *quedar(se) libre* 'to be set free') as the central member loses token frequency and the category loses types. Taken one step further, some categories show evidence of having become almost completely unproductive by the 1900's. This study shows that categories are not static when looking at them in a diachronic setting. Furthermore, the data show that these categories have longevity; many of them endure from the 1200's to the 1900's (see Chapter 5).

Other factors were examined that may have affected *quedar(se)* + ADJ as a construction. In examining ascendant and descendant trends in category productivity, this study looks at interactions with other verbs of 'becoming'. In the first century from which data was collected, the semantically similar *fincar(se)* + ADJ was prevalent in usage with adjectives that would go on to occur with *quedar(se)* + ADJ, or that were semantically related to types used with the latter. This could be partially due to the fact that the two verbs were synonymous, making it plausible for one to be used in a manner similar to the other. It is likely that, as *fincar(se)* + ADJ underwent a loss in overall frequency, *quedar(se)* + ADJ 'adopted' many of its adjectives. In the later years, from the 1600's onward, there is evidence that adjectives used in *quedar(se)* + ADJ went on to be used with other verbs, and that this correlates with the loss in overall frequency of this construction during this period. At the same time, the portion of usages of the verb *quedar(se)* to express 'becoming' increases through time almost reaching the portion of

usages to express 'becoming' in the 1800's. Chapter 4 presents this background information.

In order to gain a better understanding of frequent combinations of *quedar(se)* + ADJ, this study uses measures of relative frequency in order to identify both prefabs and instances of the construction that show evidence of conventionalization. These proposed measures, or thresholds, are used in order to support the claim that prefabs have longevity, that frequent combinations are the central members of categories, and that there is a correlation between the token frequency of the central member and the type frequency of the category to which it belongs. Furthermore, the observed correlation between token and type frequency provides evidence that categories have gradient structure. As a category becomes productive, it initially attracts types to the central member. In losing productivity, marginal types disappear as the central member remains. The manner of reaching these thresholds is described in Chapter 4 and elaborated upon in Chapter 5.

Finally, in order to compare the proposed structure of the exemplar clusters in Chapter 5 to structures based on data from a larger group of native Spanish speakers, a similarity study was conducted using a multi-dimensional scaling (MDS) analysis. Two separate analyses were executed using data from two different questionnaires that asked participants to rate the similarity of adjective pairings on a five-point scale. As with Bybee & Eddington (2006), the configurations produced by the analysis provided empirical support for the proposed structure of the exemplar clusters. The results of this multivariate analysis are presented in Chapter 6.

The models underlying this research, namely construction grammar and the exemplar model, are described in Chapter 2. Chapter 3 explains the methods for collecting data and performing the analysis. Chapter 4 provides background information necessary to the analysis including information on the overall frequency of *quedar(se)* + ADJ, thresholds pertinent to prefabs and conventionalized instances of constructions (CIC's), the interaction of *quedar(se)* + ADJ with other change-of-state expressions, and comments on the possible roots of *quedar(se)* + ADJ. Chapter 5 presents four different sets of exemplar clusters of adjectives used in *quedar(se)* + ADJ from the 1200's to the 1800's. This chapter analyzes their development in order to account for emergent trends. The results of the MDS analysis are presented in Chapter 6 and Chapter 7 is the conclusion. Sets of clusters that were not analyzed in Chapter 5 are shown in Appendix 1 and the works used for the database are given in Appendix 2.

Chapter 2: Theoretical background: the exemplar model, construction grammar, and frequency effects

In his thorough review of research supporting the exemplar model's benefits over other models of categorization, Chandler (2002: 96) succinctly concluded that, "Linguistic usage creates grammar, not the other way around." The idea that grammar emerges from usage is widely accepted and is seen as the underpinning of usage-based linguistics (Bybee 2001, 2006; Givón 1979; Hopper 1987; Hopper and Thompson 1984; Tomasello 1998, 2003). Applied to the ongoing evolution of language through time, the focus of this dissertation, the concept of emergence through usage is a central concept. One of the main things that this thesis seeks to demonstrate is the process by which prior linguistic usage affects emergent linguistic usage. This is achieved through a corpus-based investigation of adjective categories found in the change-of-state construction *quedar(se)* + ADJ over the span of several centuries. The purpose of this chapter is to describe the theoretical framework upon which this investigation has been implemented with an emphasis on showing how it is relevant to the study of diachronic data. This chapter contains a discussion of the exemplar model, its advantages over similar models and its application to a historical study. The second section discusses usage-based construction grammar and how it relates to the methodology of this study. The third section describes the role of frequency in ongoing language change.

2.1 Exemplar model

The first implementation of what is presently called the exemplar model has been attributed to Medin & Schaffer (1978: 207) who carried out a psychological study of the

categorization of geometrical shapes as realized by their participants. Their theory was developed partially in response to existing models of perception and categorization that they found to be inadequate in explaining these processes. Summing up their main point regarding categorization, they wrote that, a probe item (i.e. incoming stimuli) acts as a retrieval cue to access information associated with previously experienced stimuli that is similar to the incoming probe. This mechanism allows for reasoning by analogy in which categorization of new stimuli is based on comparison to previously experienced exemplars that are stored in memory along with other contextual and experiential information. The defining characteristics for categorization, it follows, are not defined by a subset of features (Medin & Schaffer 1978, Chandler 2002, Bybee & Eddington 2006), as proposed in the prototype model. Instead, each experience is categorized according to a comparison to every other instance of that same experience or, if it is a new experience, according to experiences perceived as somehow similar. This investigation uses the terminology used by Bybee (2001, 2006) and Bybee & Eddington (2006) in which "such representations are called EXEMPLARS and the term TOKEN is reserved for the actual instances found in usage" (Bybee & Eddington 2006: 325). In a similar vein, Bod (2006) defines a token as an instance of linguistic use and exemplar as a categorization, classification, or analysis of a token (i.e. a type). Therefore the "probe item" proposed above by Medin & Schaffer (1978) would be a token, or episodic experience, and the accumulation of stored information that is associated with the probe would be the exemplar. Applied to the present study of adjectival categories found in the construction *quedar(se)* + ADJ, exemplars correspond to adjective types. Another point that will be

developed throughout this chapter is that each experience, or token, slightly modifies the entire system (Gahl & Yu 2006, Wedel 2006).

Because it treats linguistic representation in the same way as other cognitive processes, as opposed to other models that assume that language is operated by a distinct mechanism in the brain, the exemplar model provides an ideal theoretical underpinning for linguistic representation and change. Instead of being based on abstract rules, such as generative models, the exemplar model advances the point of view that human language production and comprehension are cognitive functions based on concrete linguistic experiences that are stored in memory. This is in contrast to a formalist type of grammar in which the goal is to develop the smallest and most efficient set of rules that will generate all possible utterances in a language, and in which the speaker is deemed to have an innate knowledge of grammar (Chomsky 1965).

In exemplar theory, "each category is represented in memory by a large cloud of remembered tokens of that category" (Pierrehumbert 2001: 140). These exemplar categories are stored in a cognitive map where similar exemplars are stored close together in cognitive space and dissimilar ones are far apart. Exemplars with high token frequency have a robust representation which causes them to occupy a central position in a gradient category structure. Exemplars with lower token frequency are less robust and are proposed to occupy marginal positions in gradient category structures.

In the input, an incoming token is compared to existing exemplars, and then classified according to the most similar one; this strengthens the mental representation of the existing exemplar. If there is no similar exemplar, the probe is analogically classified according to its perceived similarity to other existing ones, thereby forming a new

exemplar. In this way, clusters of conceptually similar exemplars develop and are metaphorically represented in cognitive space as close to, or far from, other clusters according to their perceived similarity.

In order to illustrate this process, Pierrehumbert (2001: 143-44) gives a simplified hypothetical case in which the labels /I/ and /ɛ/ are possible phonological categories for an unknown vowel token that would be encountered by a supposed speaker. This hypothetical token has an f2 value that places it in an ambiguous zone in which there are exemplars of both /I/ and /ɛ/; this overlap may come from experiences of dialectal differences, speaker differences, or random variation. However, in this ambiguous zone, there are seven highly activated exemplars of /I/ and two less activated exemplars of /ɛ/. Activation here is consummate to an exemplar's representation in memory; a higher level of resting activation indicates that an exemplar has higher token frequency. In this hypothetical scenario, there are more exemplar types labeled /I/ in this range of exemplars. Thus, the unknown vowel token is assigned the label /I/ and a new exemplar is formed. However, Pierrehumbert admits, this model is highly simplified and does not take into account the various dimensions of episodic information that would be associated with such a variant.

In the case of novel usage, a category can be formed on the basis of a single linguistic event. Medin & Schaffer (1978) pointed out that memories of one experience were enough to form categories. Over time, a novel token may either 'die-out' or, if reinforced, serve as the basis for forming a new exemplar. As Chandler states (2002: 96), "Exemplar-based models imply that categories and categorization arise spontaneously when a probe enters our working memory and evokes into activation those memories that

share experiential features with the probe." Although this research operates from the point of view that exemplar categories are more permanent, given the prevalence of repetition in the human experience, Chandler's (2002) observation aptly describes novel input. Categorization occurs as a reaction to each and every token and that it occurs constantly in the human experience in a variety of cognitive modes. While exemplars tend to represent category types with numerous tokens, it is possible for an exemplar to be formed on a single event, as in Pierrehumbert's example above.

Because exemplar theory deals with different dimensions of categorization, it is logical that several dimensions of categorization are relevant to linguistic categorization including, but not limited to, semantic, phonological, morphological, situational, and pragmatic levels (Pierrehumbert 2001). In other words, these are the "experiential features" mentioned by Chandler (2002: 96). Medin & Schaffer (1978) had similar proposals in their experimental investigation of categorization; information regarding the cue, the context, and the event are all stored together. This leads to two important points.

First, Bybee & Eddington (2006: 325) observed that cognitive representation consists of several levels of abstraction. In talking about their research they used the example of *ponerse nervioso* 'to get nervous' in which a specific verbal form such as *me pongo* 'I put (pronominal)' would be stored as a set of phonetic exemplars (i.e. bilabial nasal + mid-front vowel, etc.). At a possibly higher and more abstract level of representation, all verbal inflections of *ponerse* would be stored together. At an even higher level of abstraction yet, all instances experienced by the speaker of *ponerse* used with an adjective used as a change-of-state construction would be represented. These specific verb + adjective combinations, constructions consisting of two or more words are

proposed to be represented together in memory as exemplars of the change-of-state construction *quedarse* + ADJ.

Second, the different dimensions of categorization are important because, as Bybee & Eddington (2006) stated (citing Chandler 2002), redundant or marginal features of an experience may be used as a basis for category expansion. This opens up the possibility of analogical extension based on any dimension of experience associated with that category including established exemplars, marginal exemplars, or experiential information.

In the discussion up until now, the focus has been on what happens to incoming exemplars and how they are perceived. The formation of exemplar clusters has an impact on production as well; generally this is based on categorization but proceeds in the opposite direction. Bod (2006: 295), in applying the exemplar model to syntax, states that "language *comprehension* corresponds to computing the most probable meaning given an utterance while language *production* consists of computing the most probable utterance given a meaning." In applying the exemplar model to phonological processes, Pierrehumbert (2001) describes the perception of phonetic input as a process whereby the encoded phonetic character of an incoming stimulus would locate it in parameter space; production is the same process but it moves in the opposite direction. Yet she acknowledges that, in production, the selection of one exemplar over another could be random, or there may be deeper causes, such as social and stylistic factors that might cause one exemplar to be chosen over another in certain situations. Despite these two observations, she proposes that, in most cases, the probability that a specific exemplar will be selected is proportionate to its strength of representation; this strength is based on

token frequency. Even though Pierrehumbert (2001) makes these observations based on phonology, these concepts translate plausibly to syntactic exemplars such as verb + adjective combinations. Based on the high strength of their mental representation, it is improbable that adjectives in exemplars with high token frequency, such as *solo* in *quedar(se) solo* 'to be left alone', would be produced with any other verb as an expression of 'becoming'. For example, it is not probable that a native speaker would say *me hice solo*, or *me puse solo* to mean 'I was left alone' instead of *me quedé solo*.

In the case of novel expansion (i.e. the appearance of new adjectives in *quedar(se)* + ADJ over time), the process is somewhat different. As pointed out in Bybee & Eddington (2006), redundant or marginal features could serve as the basis for the novel expansion of a category (see Chandler 2002), and it is not necessary to predict which features are chosen since they are all represented. However, Bybee & Eddington observe, conceptual clustering can give an idea of how subsequent uses of a particular construction will occur over time by showing which verb + adjective combinations are already established. Even though many factors may contribute to the production of novel forms, analogical extension based on the semantic similarity of previous adjective usage is the most plausible vehicle of category expansion.

2.1.1 The exemplar model vs. the prototype model

The exemplar model and the prototype model are both models proposed to account for categorization. Although they have many similarities, there are some fundamental differences in the way that these models account for categorization, production, and mental representation. Even thought the exemplar model may include some of the proposals of the prototype model, it is considered by many researchers to

better represent cognitive categorization (Bybee & Eddington 2006, Chandler 2002, and Skousen 2002). This section aims to describe the prototype model and to show how it relates to the exemplar model. Furthermore, this section seeks to advance the argument that the exemplar model is better at accounting for categorization and category expansion. Specifically, it provides a more plausible in account of the emergence of linguistic patterns observed in the diachronic data used for the present study.

2.1.1.1 Prototype theory

In general terms, prototype theory proposes that people learn, or acquire, a summary representation of an entire category and that they use this representation to determine category membership (Murphy 2002, Rosch 1978, Taylor 1995). The process of category learning is one of determining what is the best prototype of a given category, and categorization is a process of comparing incoming items to the prototype. In other words, for each category, there is a prototypical member that best represents the category as a whole based on a group of features that the prototype possesses. For example, according to prototype theory, a person's category of birds would be represented by an "ideal bird" Rosch (1978) that best exemplifies the characteristics typically attributed to birds. This prototype is used to categorize other animals that the person encounters. Incoming items are deemed to be classified according to how many features they have in common with the prototype. If an item has enough features in common with the prototype, it is categorized as a member of that category. Therefore, if an animal encountered by a person walks on two legs, has a beak, has feathers, and has wings, as does the prototypical bird, it must be a bird. This also takes account of birds that don't fly (chickens), that are much bigger than average (ostrich), that live in water (duck), etc.

It follows that references to prototype effects have to do with the idea that there is one prototype that represents the best, most adequate, member of a category. All other members of the category have a gradient membership in relation to the best example. Some are closer to the ideal and some are more marginal. This gradient category structure is called family resemblance and is incorporated into exemplar models of categorization, as will be discussed in section 2.1.1.2 below.

Even though proponents of prototype theories might not deny that episodic memory plays a part in categorization, they claim that episodic information is stored separately from category information (Murphy 2002: 20). Taylor (1995) states that episodic events do have an impact in prototype representation in that categories are created by extracting features from specific events but that these exemplars are not stored in memory. Contrary to this, there is widespread agreement among researchers who implement the exemplar model that specific instances of experience are part of stored category information (Chandler 2002, Bod 2006, Pierrehumbert 2001, Wedel 2006). However, according to the prototype theory, these experiences upon which exemplar categories are formed are not stored in long term memory. This creates further problems when researching language change because it does not allow for categories to undergo slight alterations with each new experience as the exemplar model proposes.

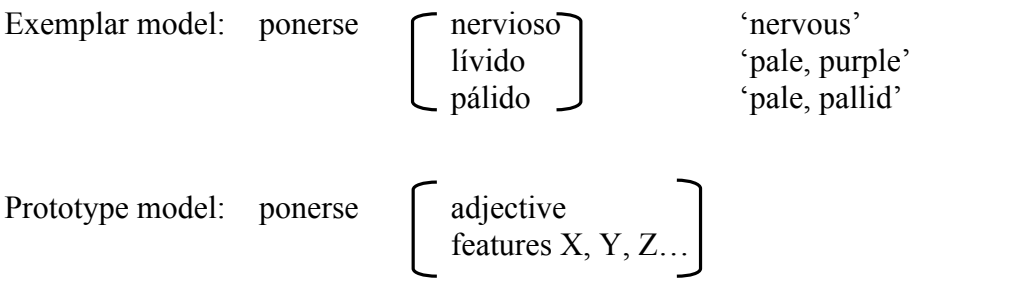
2.1.1.2 Advantages of the exemplar model

In accounting for the phenomenon of categorization, especially of linguistic input, researchers have claimed that the exemplar model not only outperforms the prototype model in psycholinguistic experiments (Chandler 2002, Skousen 2002) and psychological experiments (Nosofsky 1998), but is also much more plausible in representing language

change (Bybee & Eddington 2006). Exemplar theory explicitly rejects the notion that there is a prototype that somehow embodies an entire category. Instead, a person's concept of a category is based on the set of all experienced episodes with members of that category. The ideal bird is irrelevant; it is the set of all previous experiences with birds that are used to categorize new encounters with birds. Any notion of an "ideal bird" would be due to the fact that a person will have had a higher frequency of experiences with certain birds. Over time, a person's experience with birds may change and each new experience with them slightly alters the previously formed categories of birds. Therefore, even though there may be more centrally categorized birds and more marginally categorized ones, circumstances may arise that would cause a more marginal member of the category to gain central status.

In linguistic representation, as with other types of mental representation, the exemplar model and the prototype model have different ways of depicting category formation. Figure 2 illustrates how Bybee & Eddington (2006: 326) demonstrate this difference with the change of state construction *ponerse* + ADJ.

Figure 2. Exemplar model vs. prototype model in representing *ponerse* + ADJ, from Bybee & Eddington (2006: 326).



Whereas the exemplar model lists adjectives according to semantic similarity, the prototype model "would not list the actual adjectives, but would rather extract features

from them" (Bybee & Eddington 2006: 326). These features would then be compared to the features present in the prototypical adjective, or to the feature list that governs a particular category of adjectives. In the exemplar model, speakers choose verb + adjective combinations based on comparison to exemplar categories formed by prior linguistic experience and/or analogy to prior experiences. Bybee & Eddington (2006: 326) explain that not only is semantic similarity to previous usage important, but that frequency also plays a key role in categorization of adjectives. Frequency effects support an exemplar model of representation and, also, allow for the observation that there are some prototype effects in this model of representation. Specifically, exemplar categories demonstrate gradient category structure in which some members, exemplars with high token frequency, are more central than others. Members of these categories could be conceived of as prototypical, especially since they tend to display most of the features common in other members (Bybee & Eddington 2006). There are a number of researchers who illustrate how the exemplar model explains many of these prototype effects (see: Hintzman 1986, Logan 1988, Nosofsky 1986, 1992, and Nosofsky & Palmeri 1997).

In another investigation, Nosofsky (1988) studies the relationship between the interactive roles of frequency and similarity in determining graded category structure. His two experiments support the hypothesis that both frequency and similarity are a part of people's representation of categories and that both of these variables affect graded category structure. He found that classification accuracy and typicality ratings would increase for exemplars that were experienced with high frequency and for exemplars that were similar to high frequency ones in the target category. He claims that his research is

evidence that categories are learned through storing individual experiences in memory and that a frequency-sensitive exemplar model is better at accounting for graded category structure than a frequency-sensitive prototype model.

Also, like prototype categorization, there is 'family resemblance' in the exemplar model in which marginal members may share characteristics with the central members but not necessarily with one another. In these cases, it is the central member that links marginal members, as will be demonstrated in Chapter 5:. As mentioned above, there is graded membership in which some members are more central and some are more marginal. However the boundaries of inclusion or exclusion between possible members of a category are not discrete (Bybee & Eddington 2006, Lakoff 1987).

There are other reasons why the exemplar model is deemed by many researchers to be better at accounting for linguistic categorization. One of the key arguments is that it recognizes that marginal or redundant features can serve as the basis for category expansion (Bybee & Eddington 2006, Chandler 2002). Since the exemplar model proposes that all episodes of experience are stored in memory with all of the features of a category, it is unnecessary to predict beforehand which feature would emerge in leading the expansion of a category. Even though high-frequency exemplars may have the effect of making category formations more centralized, marginal members may serve as models of analogical extension that expand a category into new domains.

2.1.2 The exemplar model and diachronic data

Wedel (2006: 269), in his diachronic study using the exemplar model, observes that language change must be conceived as resulting from "two nested models: a smaller scale driven by language use by speakers, and the longer scale cycle of transmission

through generations". The exemplar model is ideal in representing diachronic linguistic change because a speaker's new experiences are proposed to modify the existing cognitive representation and previously formed categories may undergo a shift (Bybee & Eddington 2006, Gahl & Yu 2006, Wedel 2006).

In applying the exemplar model to account for language change, Wedel (2006) simulates three potential pathways of change based on the theory of evolution: pruning of lines of inheritance (also known as genetic drift), blending inheritance, and natural selection. In his paper, he states that the processes of pruning lines of inheritance and natural selection may explain evolutionary change in language. He justifies his stance by noting that the mechanism of evolution is independent of biology and can apply to any phenomenon by which there is an ongoing population of variants, such as the different phonological exemplars present in an exemplar category. Even though independent from biology, he goes on to state that patterns of evolution in biological systems may provide potential models by which to study language change. Viewed this way, every exemplar can trace its ancestry back to a line of descent. Pruning lines of descent inheritance refers to the process by which some exemplars with low activation decay in memory; when an exemplar decays significantly before being produced, its line of descent is truncated, or pruned. Natural selection refers to the process by which intervening factors affect the likelihood that a given exemplar will be transmitted to the production / perception loop thereby changing the frequency of this variant. To illustrate possible intervening factors he refers to contextual biases toward lenition (reviewed in Bybee 2001) or hyperarticulation which may serve to increase the frequency of some forms and cause shifts in the category. While natural selection can result in an increase or decrease in the

number of variants, pruning always results in a decrease. On the other hand, blending inheritance is possibly the vehicle that helps maintain category cohesion over time. This process stems from the proposition that there is a tendency for the perception to be warped away from the extremes of a category and be biased toward the center in which the population of exemplars is more dense (see Wedel 2006 for more detail).

Although Israel (1996) doesn't explicitly use the exemplar model in his diachronic analysis, his research is relevant the present study. It is a usage-based diachronic study that tracks the pathways of analogical extension of the *way*-construction over several centuries. The *way*-construction consists of a fixed element, the noun *way* with a genitive NP, and an 'open slot' consisting of a verb and a frequently appearing PP as in example (22) below.

(22) *Popeye fought his way through the crowd.*

As Israel (1996: 220) points out, the usage tends to "cluster around certain narrowly defined semantic verb classes." Other fighting verbs, such as *force, claw, elbow, knee*, and *push* are highly attested in his data and could easily replace *fought* in the example above. Israel focuses on the development of two general categories over time calling them the means thread and the manner thread. These categories become more productive and acquire new verb types in the span of his data. Through analogical extension, the manner thread went from including common verbs of motion (e.g. *go, ride, run, pursue, wend*, and *pass*) and came to include more diverse predicates that encoded a manner of motion (e.g. *plod, scramble, thread, crash* and *sing*). In its early forms, the means thread was used with verbs of path clearing and creation (e.g. *pave, smooth, cut, furrow out,* and *force out*). Out of this thread emerged a category of "fighting" (e.g. *battle, push, struggle,*

beat, and *shoot*). In its later stages, the means thread used with verbs that encode increasingly indirect means of enabling motion (e.g. *blot, write, plumb, spell, smirk,* and *knave*). Over time, he notices that there is overlap between the two over-arching categories formed by the two threads observing that the categories formed by the verbal predicates are often so far removed from the original thread that it is hard to say if a particular verb type belongs to one thread or another.

There are some important similarities between the observations that Israel (1996) makes and the tendencies that are observed in the present investigation of *quedar(se)* + ADJ. Israel notes that early established usages are maintained throughout his data and that one of the remarkable things about this diachronic evolution is the consistency of usage over time. The role of analogy is a central factor in the emergence of novel usages as these are derived through analogical extension based on earlier usages. He also states that speakers must track specific verbs and their frequency as they organize linguistic knowledge. This leads to an overall scenario in which new forms entering the construction tend to cluster around certain "well defined semantic prototypes" (Israel 1996: 222).

Both Wedel (2006) and Israel (1996) show the importance of diachronic models of categorization in the understanding of present linguistic usage. There is constant innovation as language is transmitted from speaker to speaker, and from generation to generation. The exemplar model is perfect for representing this process because it shows how a speaker forms categories through linguistic usage and, in turn, uses these categories to interpret and produce novel tokens. The constant exposure to new linguistic experiences results in gradual updates of categories that can be documented throughout a

speaker's life and culminate in even more notable category changes when tracked through many generations.

2.2 Construction grammar

The idea of construction grammar (Croft 2001; Goldberg 1995, 2003, 2006) is also necessary to the present study of *quedar(se)* + ADJ. Although construction grammar is a model of linguistic representation separate from the exemplar model, the two share many complementary ideas. This section provides a general discussion of construction grammar and seeks to establish its relevance to the present investigation.

2.2.1 Construction grammar and the exemplar model

In her book on constructions and construction grammar, Goldberg (1995) continually reinforces the notion that there are clusters of categories formed by the variable types in a construction and that these clusters are formed based on some dimension of similarity. She acknowledges that novel uses are based on analogy to previous uses. By including notions derived from frame semantics (Fillmore 1977) she depicts a model of experiential representation similar to that of the exemplar model. Basically, experiential information, including world and cultural knowledge, are associated with a specific construction along with the linguistic information.

Especially relevant to the study at hand, Goldberg (1995) depicts the classification of verbs with 'bubbles' in a similar manner to the present study and its direct predecessors (Bybee & Eddington 2006, Wilson 2009). She explains that frequency affects classification and that similar items are stored close together. She also makes the proposal that clusters with high type frequency will have a higher level of productivity than clusters with low type frequency, a frequency effect observed also in Bybee &

Thompson (2000). Furthermore, Travis & Silveira (to appear) demonstrated that high type frequency contributed to the productivity, and consequential spreading, of an emergent first-person plural marker in Brazilian Portuguese. Other scholars have also noted an association between high type frequency, productivity, and extension in use (see Baayen 1993, Bybee 1985, Smith 2001, Tottie 1991, Zuraw 2003).

Croft (2007 to appear: 6-7) points out that applying the exemplar model to constructions may pose unique problems not only because the model is relatively young and has been typically applied to phonology, but also because there are added complexities that arise in applying it to form-meaning pairings. For example, it isn't always clear what constitutes a similar semantic situation as it is in phonetics. It's also harder to obtain a large number of occurrences in a way that is possible in a phonetic analysis. He goes on to state, however, that some very important elements of the exemplar model can be found in construction grammar. One of the underlying purposes of this thesis is to show how the exemplar model, applied to diachronic data, can account for the emergence and development of semantic categories that stem from the adjective slot of the construction *quedar(se)* + ADJ.

2.2.2 Construction grammar: basic principles

One of the main assumptions of construction grammar is that, as Goldberg (1995: 4) puts it, "constructions are taken to be the basic units of language" and they are highly routinized even though they are readily extended to new contexts in principled ways. As a general definition of these pervasive "basic units of language", Goldberg (2003: 219) defines constructions as "stored pairings of form and function, including morphemes, words, idioms, partially lexically filled and fully general linguistic patterns." Goldberg

(2006) goes so far as to state that the sum of our grammatical knowledge is the sum of our knowledge of constructions; in other words, everything in language is a construction. Similarly, Croft (2001, 2003) views constructions as the basic, primitive units of syntactic representation. These are mapped onto a syntactic space in the way that phonetic variables are mapped onto phonological space: through usage, experience, and reinforcement (Croft 2007 to appear).

Another central theme of construction grammar is that it is a usage-based account of language. Goldberg (2006) explains that information about the usage of constructions, such as frequency, patterns of usage, and specific instances of use, are stored in memory. Even though there is a great deal of information regarding specific instances of use (i.e. tokens), categories are formed based on meaningful relationships between the types in a category. The similarity between types based on these more general ideas of a category is what leads to category cohesion.

The idea that constructions are considered to be core grammatical units also indicates the marked difference between construction grammar and top-down generative grammars that seek to describe a language through transformative rules (Chomsky 1957). Generative grammars of this ilk propose that usage is the result of some sort of transformation of a core unit as it interacts with a higher linguistic dimension (i.e. lexical roots interacting with morphological transformations in verbs). In formulating these rules, researchers of generative grammars assume that there is a strict division between syntax, phonology, morphology, and the lexicon and that each of these components describes only one aspect of the characteristics of a sentence. Research in construction grammar tends to assume the opposite. Many proponents of construction grammar (Goldberg 1995,

2003, 2006; Croft 2001; Langacker 1987; Barlow & Kemmer 2000, Noonan 1998) state that there is no strict division between the lexicon and syntax, and that there is no strict division between semantics and pragmatics either. Since all dimensions of language are viewed as comprising equal parts of a speaker's linguistic knowledge, any of these may contribute to the usage of a particular construction in a particular way. Langacker (1987), in describing cognitive grammar, itself a usage-based model that deals with constructions, proposes that there is a grammar / lexicon continuum and that the cognitive representation of language involves categorization on many levels of abstraction. One of these levels of abstraction may contain specific episodic information about the usage of a construction.

Goldberg (2006), in discussing construction grammar, also indicates that categories associated with constructions may extend into new territory. Instead of analogical extension, Goldberg uses the term *generalization* to explain novel utterances. Speakers create novel uses of a construction by generalizing constructions beyond their originally experienced contexts into new ones based on accumulated notions that speakers have regarding overall categories. As Goldberg (2006: 229) writes in the conclusion of her book, "The generalizations of language, like generalizations in other cognitive domains, are formed on the basis of instance-based knowledge that is retained."

Even though Goldberg (1995, 2003, 2006) proposes that even morphemes and words are constructions, this study is concerned with the nature of multi-word sequences. Regarding formulaic strings of two or more words, there are two extremes: on one end, fixed idiomatic strings which are semantically obscure and syntactically immutable such as *by and large*, and, on the other end, flexible strings that have both fixed elements and

open slots that allow for the substitution of a variety of items, usually of the same grammatical category (Wray 2000). Bybee & Cacoullos (2009) state that there is no discrete division between these, but a continuum in which fixed expressions and productive constructions form the two poles. Referring back to Goldberg's (1995) definition of constructions, immutable strings of two or more words would be the idioms and the "partially lexically filled linguistic patterns" would be general constructions such as the *way* construction and *quedar(se)* + ADJ. The *way* construction allows for the substitution of a variety of verbs and prepositional phrases that pertain to certain semantic classes. Similarly, the construction *quedar(se)* + ADJ allows for the substitution of a large gamut of adjectives that belong to categories of semantically similar adjectives.

The construction under consideration, *quedar(se)* + ADJ, as with many constructions, is a form-meaning pairings that must be analyzed not only by the meaning of the individual parts, but as a whole unit. While the adjective expresses senses that are congruent with its broader usage outside the construction, the verb does not necessarily do so; what the verb typically means may not always play a determinant role in the meaning conveyed through their usage as verbs of 'becoming' (Eddington 1999, 2002, Bybee & Eddington 2006). It is more relevant to "ask not so much what the verb means, but what the overall construction means" (Bybee & Eddington 2006: 328). At the same time, highly routinized constructions, such as prefabs, may retain some of the older meaning or distribution of the original construction (Bybee & Cacoullos 2008). As Lakoff (1987: 465) pointed out, "grammatical constructions in general are holistic, that is, that the meaning of the whole construction is motivated by the meanings of the parts, but is not computable from them".

2.2.3 Prefabs

The concept of prefabs is important to the present study since one of the claims made here is that prefabs have longevity. Prefabs and highly frequent types have other important roles in the development of categories; central members of categories tend to be prefabs or types with relatively high token frequency, and there is a correlation between the token frequency of the central member and the type frequency of the category to which it belongs.

Erman & Warren (2000: 31) define prefabs as "a combination of at least two words favored by native speakers in preference to an alternative combination which could have been equivalent had there been no conventionalization." Yet, a categorical identification of prefabs, as opposed to general constructions, has posed some problems for researchers. Erman & Warren (2000) in their study of prefabs acknowledged that it is difficult to categorically identify prefabs for two reasons: first, what is a prefab to some members of a speech community may not be one to all members, and, second, prefabs are easily overlooked. Nevertheless, using a conservative definition of prefabs, they determined that approximately 55% of a text will consist of prefabricated sequences and they note that less conservative measures may result in even higher numbers.

Some researchers have identified prefabs on the basis of the relative frequency as compared with the usage of related forms (Bybee & Eddington 2006, Bybee 2006, and Wilson 2009 for Spanish verb + adjective expressions of 'becoming', Bybee 2003 for verbs used with OE *can*, Travis & Silveira To appear for Portuguese 1p subject + verb combinations). Implementing ideas of relative frequency from Hay (2001) and Torres Cacoullos (2000), Bybee & Cacoullos (2009: 204) offered an operational definition of prefabs in their study of "auxiliary" + gerund constructions in Spanish: "combinations

making up 2% or more of the corresponding "auxiliary" data and 50% or more of the corresponding gerund data". For example, a combination such as *estar aguardando* 'to be waiting' would be a prefab based on the fact that this combination represents 2% (14/672) of all occurrences of *estar* and 93% (14/15) of all occurrences of *aguardando* in their data. In the present investigation, Section 4.2.1.1 (Chapter 4:) offers an operational criterion for identifying combinations of *quedar(se)* + ADJ that are prefabs using data from the *Corpus del español* (Davies 2002-): (a) there is no intervening material between the verb and the adjective, (b) the adjective accounts for at least 1% of all occurrences of that adjective in Davies' (2002-) data, and (c) the combination accounts for at least 2% of all occurrences of *quedar(se)* + ADJ in the same data. See the aforementioned section of Chapter 4 for a more detailed discussion of issues regarding prefab thresholds.

One of the proposals of this study is that central members of exemplar categories with high token frequency are often prefabs[2]. This claim is echoed in other usage-based work on formulaic language (Bybee 2006, Bybee & Eddington 2006, Bybee & Torres Cacoullos 2009). This has important consequences for the development of the general construction from which the prefab may have emerged. In their study of the grammaticization of Middle English *can* and Spanish "auxiliary" + gerund combinations, Bybee & Torres Cacoullos (2009) make the case that prefabs advance the productivity of the general construction; by being the central member of semantic subclasses of a construction, they attract more lexical types into these subclasses. Productivity is also fed by the fact that prefabs lead in the semantic bleaching of a construction. This contributes to the development of an emergent meaning of the construction that is independent from

[2] For a discussion on the criteria used in determining the central member of a category, see Chapter 3:, Section 3.4, and Chapter 5:, Section 5.1.

the original meaning of the individual elements. The findings of Bybee & Torres Cacoullos (2009) relate to the analysis of *quedar(se)* + ADJ (Ch. 5) in which there is a correlation between the token frequency of the central member and the type frequency of the category to which it belongs.

2.3 Frequency and language change

Frequency has been mentioned in this thesis as playing an important role in the exemplar model of categorization. Frequently occurring forms have more robust representation in exemplar categories and token frequency is a criterion for identifying central members in exemplar clusters. Furthermore, combinations that demonstrate a significantly higher relative frequency than others are often considered to be prefabs (Bybee & Torres Cacoullos 2008). This section seeks to consolidate and further discuss the importance of frequency as it affects language change and linguistic categorization.

Directly related to the study at hand, Bybee & Eddington (2006: 349-52) confirmed in their acceptability experiment (see Chapter 1:, Section 1.2.2.2) that frequency plays a role in acceptability judgments of verb + adjective combinations involving the two verbs of 'becoming', *ponerse* and *quedarse*; frequently occurring exemplars are more acceptable than infrequent ones. Also, exemplars that are semantically similar to frequent ones were judged as more acceptable than those that were not similar. This suggests that novel forms will be judged as acceptable if they bear similarity to previously experienced patterns of usage than if they do not.

Shifts in exemplar clusters could be explained in terms of shifts in frequency and probability. As a form gains in token frequency it may form its own new exemplar cluster (Bybee 2002b). In her study on probability and language change, Zuraw (2003) offers

some insight that is relevant to a usage-based model of language change. She writes that speaker interaction is the locus of language change and that each interaction modifies the grammar and lexicon slightly. These modifications result in incremental changes because they affect the probability that a structure will be produced with a higher or lower frequency. As a certain form is experienced at a higher frequency, the representation grows in strength, and the probability that it will be produced increases. A similar but opposite process affects forms of diminishing frequency. Frequency shifts also have an effect on the re-analysis of a construction that might even cause it to function in a new grammatical category.

High token frequency can both promote and slow change. Bybee & Thompson (2000) refer to this as the 'Reducing' and 'Conserving' Effect respectively. Travis & Silveira (To appear) documented these two frequency effects in their study of first person plural subject pronouns in Brazilian Portuguese: the older pronoun *nós* 'we' and the emergent one *a gente* that developed from an NP meaning 'the people'. A rise in token frequency is one of the chief factors that promoted the semantic bleaching of *a gente* leading it to be analyzed as a 1p pronoun as it extended into usage with more verbal types. However, the Conserving Effect was evident only in cases of extremely high frequency with the older pronoun *nós*. They found that the retention of *nós* with the most frequent hortatives (e.g. *digamos* / *vamos dizer* 'let's say', *vamos ver* 'let's see', and *vamos* 'come on') was so conventionalized that these combinations were resistant to being used with *a gente.*

There is also a relationship between productivity and type frequency. Goldberg (1995, 2006) and Bybee & Thompson (2000) discuss the way in which type frequency

plays a key role in a construction's degree of productivity. Expanding upon notions from Bybee (1985), Bybee & Thompson (2000) state that a higher type frequency will preserve the exchangeability of the open slot of a construction. Even if this a certain combination becomes highly entrenched, as exemplified with *nós* and *a gente* in Brazilian Portuguese above (Travis & Silveira To appear), Bybee & Thompson (2000) state that these highly entrenched combinations may continue to be used even alongside more productive usages. As Goldberg (2006: 99) puts it: "A pattern is considered extendable by learners only if they have witnessed the pattern being extended". She goes on to say that the extension of types in a given category will be based on semantic similarity to established types in that category. This will be demonstrated to be the case with *quedar(se)* + ADJ.

2.4 Summary

This section has described the exemplar model, an instance-based model of categorization (Section 2.1), and has argued that it is an ideal model at representing linguistic change (Section 2.1.1.2), especially regarding diachronic data, such as the development of *quedar(se)* + ADJ (2.1.2). Speakers develop categories associated with linguistic material based on their experiences with that form including information regarding frequency, context of usage in discourse, and social context. While all of these factors play a role in the choices that speakers make, the stance that this study takes is that previous usage affects latter usage: speakers tend to produce previously experienced forms, and novel usage is largely based on analogical extension to established forms.

Because *quedar(se)* + ADJ is a construction, this chapter presented some of basic principles of construction grammar (Section 2.2). Construction grammar is a usage-based

model that complements the exemplar model (Section 2.2.1) and is important in considering frequent sequences such as prefabs (2.2.3).

Finally, this chapter discussed frequency of usage and its relevance to language change (Section 2.3). The token frequency of a combination is a criterion for identifying it as a central member of an exemplar category and can affect the probability that a particular combination will be produced (Zuraw 2003). Token frequency can both promote and hinder language change (Bybee & Thompson 2000, Travis & Silveira To appear). Researchers (Goldberg 1995, 2006, and Bybee & Thompson 2000) have found that a correlation between type frequency and productivity. Furthermore, this study, as does Bybee & Torres Cacoullos (2009), finds that there is a correlation between the token frequency of the central category member and the type frequency of the category of which it is a member.

Chapter 3: Data and Methods

This chapter provides a detailed account of the corpus, the database, the extraction of adequate tokens, and the methodology used in applying the exemplar model to account for category emergence and ongoing representation.

3.1 Data source

The corpus consists of a total of 5,903,657 words and comes from a variety of sources both online and from electronic concordances. Table 1 shows the number of words in the corpus per century. A complete list of all works used in the corpus is presented in Appendix 2. Works were only included in the corpus if they contained at least one token of *quedar(se)* + ADJ used as an expression of becoming; all others were omitted. Also, only narrative or narrative-like works were chosen. The figures in Table 1 reflect the total number of words in the corpus from entire works.

Table 1. Total number of words in corpus and number of works consulted by century.

	Number of words	Number of works consulted
1200's	1,031,755	4
1300's	789,426	4
1400's	1,658,165	24
1500's	558,475	24
1600's	463,529	27
1700's	435,055	11
1800's	967,252	25
Total:	5,903,657	119

The 1500's, 1600's, and 1700's show a notably lower overall word count than the preceding and following centuries. This is due to a rise in the overall frequency of the change-of-state construction *quedar(se)* + ADJ. Because of this, it was not necessary to consult as many texts in order to obtain the target number of tokens per century from the

1500's to the 1700's (this 'target number' is discussed below). The fact that the 1300's has the lowest number of words simply reflects the lack of sources reliably attributed to this century in an attempt to build the corpus. Type, token, and overall frequency play an important part in the analysis of the data and will be discussed in more detail in the following chapters.

In pilot studies for the current project (Wilson 2006, 2007, 2009), the *Corpus del español* (Davies 2002-) was used as the chief source of data. In the present study, Davies' (2002-) data was used in two scenarios. First, because of a scarcity of other sources, it was necessary to include data from the Davies (2002) corpus for the 1200's. In these cases, dates were cross-referenced with two bibliographies (Faulhaber et. al. 1984, 2002) and frequency measures were calculated based on the number of words in the works from which the tokens came in Davies' (2002-) data. Second, it was used to fill in gaps in the data where a type was conspicuously absent in these data given that it appeared robustly in surrounding centuries. In these cases, Davies' (2002-) data was used to confirm a type's endurance despite the conspicuous absence (see Section 5.2.2.3, 5.2.2.6, 5.2.3.1, 5.2.3.5, 5.2.3.6, and 5.2.3.7).

Since the present study aims to track language change, historical authenticity was the single most important criterion in choosing texts for the corpus. In the earlier centuries (especially from 1200's to the 1400's) it is often difficult to accurately assign a work to a specific century. In considering a work, be it from O'Neill (1999), Davies (2002), or another source, I cross-referenced the proposed date with SPDT (specific production date) and the OPDT (original production date) given in Faulhaber et. al. (1984). If there was a discrepancy in century, the online bibliography (Faulhaber et. al.

2002) was consulted in order to see if the discrepancy could be resolved; this source provides more detailed information about the author and scribes who could have possibly updated the work. If the discrepancy was too great to resolve by consulting these two sources, the work was not included in the corpus.

As mentioned above, for a work to be considered for inclusion, it had to be a narrative or narrative-like text from the Iberian Peninsula including letters reporting events, novels, plays, narratives of events, and epic poems. By focusing on narratives, it provided a data source that was as homogenous as possible and that was closest to speech. Dictionaries, grammars, legal documents, recipe books, and other works that reflect genres that are markedly different from narratives were not used. The central point of this study is to provide a plausible theoretical model of ongoing linguistic representation of exemplar categories formed by speakers of the adjectives in *quedar(se)* + ADJ. In light of this, narratives are considered in this study to be the most authentic data for representing what speakers may have said and heard on a daily basis. Also, plays and epic poems are a good approximation of what speakers might actually do because they were often performed before a crowd and would have had to reflect a register that people would understand.

In order to further narrow the scope of the data and to further attempt to bring it in line with authentic speaker behavior, only texts from Spain were chosen. Since this study aims to show how language evolved through continuity, it was necessary to extract data from a general region in which this process would occur over seven centuries. Even though there has been considerable dialectal variation within Spain, from medieval to modern times (Lapesa 1980, Lathrop 1996, Penny 2000, Schwegler & Kempff 2007), this

at least narrows down the possibilities somewhat. This is more relevant as the Spanish Empire expanded into the New World, giving way to even more variation as many new dialects evolved.

3.2 Data extraction

Once texts were selected according to the criteria described above, they were loaded in their entirety into *ConcApp* (Greaves 1993-2003), a concordance program, and all instances of *quedar(se)* were extracted. Once a list was generated, all occurrences of *quedar(se)* with an adjective or a prepositional phrase, and an animate subject, were entered into an database for the corresponding century along with the surrounding context. From this list, all tokens were analyzed to indicate whether or not they indicated a change of state; all tokens that did not were discarded. The goal was to obtain a sample of *quedar(se)* + ADJ that was representative of the usage of adjectives in this change-of-state construction on a century by century basis. Coding was done by the researcher, but all ambiguous tokens were also coded by a second analyst in order to determine if they were truly ambiguous, or if it was possible to classify them as expressing a change of state or not[3]. However, some of the earlier tokens, especially in the combination *quedar(se) solo* naturally showed ambiguity, as in Example (23) below. In fact, this inherent ambiguity may have contributed to the re-analysis of *quedar(se)* as a change of state verb; these instances were included in the database and figure into the subsequent research. Basically, by remaining in the temple, Jesus undergoes a change of state in which he is "left alone" by the movement of other agents.

[3] Thanks to Anthony Cárdenas for help in clarifying the meaning of many medieval tokens.

(23) *E Asy commo yvan leyendo asy se yvan saliendo vno a vno Ca non se fallaua njnguno dellos syn pecado & quedo ihesu xpisto solo enel tenplo conla muger & tornose aella.*
'And as they went reading they went leaving one by one because there was not one of them found without sin and Jesus Christ was left alone / remained alone in the temple with the woman and he turned to her.' (*Castigos y documentos para bien vivir*. 13th c. Sancho IV. O'Neill 1999)

However, the majority of tokens analyzed could be clearly attributed to express 'to remain', as in Example (24) (and were therefore excluded) or 'to become' as in Example (25):

(24) *E despidió para volverse a su casa con tan tierno sentimiento, que por no verle verter tiernas lágrimas, me aparté de él sin hablarle, quedando con mis nuevos amos.*
'And he said goodbye to return to his house with such tender feelings, that to not see his tender tears falling, I separated myself from him without speaking, remaining with my new masters.' (*La esclava de su amante*. 17th ctry. María de Zayas y Sotomayor. Biblioteca Virtual Miguel de Cervantes)

(25) *Favoreció el Señor tan ilustre intento, no permitiendo que muriesse, ni dando lugar a que quedasse infamada. Quedó con fama y con vida, y sus acusadores, sin vida y sin fama.*
'*El Señor* favored such an illustrious attempt, not permitting that she die nor allowing her to become dishonored. She came out with fame and with life, and her accusers without life and without fame.' (*Fructus Sanctorum y Quinta Parte del Flos Sanctorum*. 16th c. Alonso de Villegas. BVMC.)

Both of these examples show *quedar(se)* used with a prepositional phrase involving *con* 'with', and (25) also has a prepositional phrase with *sin* 'without' in addition to a more common verb + adjective combination (i.e. *quedasse infamada* 'to become dishonored'). However, in (24) the verb *quedar(se)* means 'to remain' and the prepositional phrase is used to indicate where the subject stayed ("with my new masters"). Example (25) undoubtedly contains tokens of the change-of-state construction *quedar(se)* + ADJ and the tokens with the prepositional phrase have an adjectival function. Also, the collection of tokens using the prepositional phrases in example (25) is

considered a ‘multiple’; there is one verb but multiple types filling the adjective slot. Each adjective (prepositional phrase, in this case) of a ‘multiple’ was entered into the database separately, even if the verb refers to different grammatical subjects (i.e. ‘she’ the subject of *quedó* and *sus acusadores* ‘her accusers’). Consequently, the above ‘multiple’ would have constituted four adjectival tokens: *con fama* ‘with fame’, *con vida* ‘with life’, *sin vida* ‘without life’, and *sin fama* ‘without fame’. These function identically to the ‘singles’ in expressing a change of state and also may indicate perceived semantic similarity on the part of the speaker. Furthermore, prepositional phrases such as *sin fama* and *con fama* were included in the data, as was practiced in Bybee & Eddington (2006), because the prepositional phrase plays an adjectival role.

The target was 250 tokens per century (after exclusions) without omitting any tokens that appeared in the works used. This number was chosen because it was deemed to be sufficiently high to capture trends of adjective use in each century. In practical terms this meant that for each century there would be somewhat more than the 250 token minimum; I selected enough works to give at least 250 tokens, but extracted all tokens in those works which rendered over 250 for most centuries. In the early centuries, from the 1200’s to the 1400’s, the numbers are lower than 250 because of two factors: lack of available texts and a low overall token frequency of the construction *quedar(se)* + ADJ.

Table 2 shows the total number of tokens extracted for analysis per century, the number of types that these produced, and the standardized overall frequency per 10,000 words. In this table, we see that, despite the fact that the 1200’s and the 1400’s have more words per century than any other, there are fewer than 250 tokens. In the 1200’s, this is due to the fact that *quedar(se)* + ADJ had a low standardized frequency; the tokens

garnered came from a higher number of words than in any other century. Since the standardized frequency went up from 0.42 occurrences per 10k words in the 1200's to 0.96 in the 1300's, a lower number of words yielded a higher number of tokens. As with the 1200's, there was a general lack of works that could reliably be attributed to this century, but the reliable ones produced 76 tokens. In the 1400's, the number of words in the database increases as more texts were available; 24 reliably dated texts totaling 1,658,165 words produced 196 tokens. There are more tokens in these works as the standardized frequency of *quedar(se)* + ADJ increased to 1.18 occurrences per 10k words.

Table 2. Total number of tokens, words, and works consulted in database by century.

	Number of Tokens	Number of Types	Frequency: # per 10,000
1200's	43	29	0.42
1300's	76	40	0.96
1400's	196	121	1.18
1500's	271	153	4.85
1600's	280	171	6.00
1700's	253	135	5.82
1800's	257	141	2.64
Total:	1,374		

All in all, the extraction process yielded a total of 1,374 tokens of the change-of-state construction *quedar(se)* + ADJ that were extracted from 119 works in their entirety. These were categorized into exemplar clusters by century, a process which is discussed in the following section.

3.3 Analysis methods: applying the exemplar model to a semantic analysis

From the database of combinations of *quedar(se)* + ADJ, a semantic analysis of the adjective types was conducted in order to identify the broader exemplar clusters (*solo*

'alone', *libre* 'free' etc.) for each century. This similarity of the types proposed to belong to these clusters was then tested through a multidimensional scaling analysis (Ch. 6). The organization of these clusters is represented in the analysis in figures, such as Figure 3 in section 3.4 below. These patterns of organization not only depict a possible scenario for the aggregate mental representation of adjective types by speakers, but also show how these patterns of category organization connect from one century to another. This section describes the methods used to this end.

One point of reference for the current study are the categories of adjectives from Bybee & Eddington's (2006) 1900's data. The exemplar clusters that they proposed were based on a native speaker's grouping of these adjectives. They later performed two experiments involving larger groups of native speakers that provided remarkable confirmation of the structure of their proposed exemplar clusters: a similarity experiment and an acceptability experiment. Because of the rigorous steps taken to put their initial analysis to the test, and the verification of this analysis by the subsequent experiments, Bybee & Eddington (2006) is taken as a point of departure for grouping adjectives appearing in *quedar(se)* + ADJ into clusters in this study.

In cases where their analysis did not provide direct precedent as to similarities among groups of adjectives, I used my own intuition as a native speaker of Spanish[4]

3.4 The process

This section provides a figure produced in the analysis of the data in order to demonstrate how clusters were formed from the database generated. Methods of semantic analysis using the texts in the database are described. Also, the textual notation is

[4] I grew up in New Mexico as a native heritage speaker of Spanish but was English dominant.

discussed so that the reader will understand the relevance of font sizes, arrows, and other symbols used.

The first step was to determine which adjectival exemplars were the central members. As described above, exemplar theory is a usage-based model in which frequency of usage is highly important. One of the primary effects of frequency is that it strengthens representation and the member of a category with the highest token frequency is deemed to be the central member[5] (Bybee 2002a, 2006, Bybee & Eddington 2006). So after determining which adjectives were the most plausible central members based on token frequency, others were grouped with them based on semantic similarity. Exemplars considered to be more semantically similar to central members were placed more closely to it in this hypothetical conceptual space. Less similar or marginal members were placed farther from the central member.[6] Aside from using Bybee & Eddington's (2006) approach to representing categories, this method of category presentation is a way of illustrating Pierrehumbert's (2001) notion that categories are represented as clouds of individual tokens located in conceptual space as close to or far from other tokens based on similarity. Similarly, in her research on constructions, Goldberg (1995) portrays categories of verbs with 'bubbles' in a similar manner to the present study.

[5] Except in cases of autonomy in which a combination is idiomatic to the point of being independent from a catetory.

[6] Although it is standard practice to put foreign languages in italics, the figures of the exemplar clusters present adjective types in a non-italicized font in order to show consistency with Bybee & Eddington's (2006) notation and for the sake of maximizing space in the figures

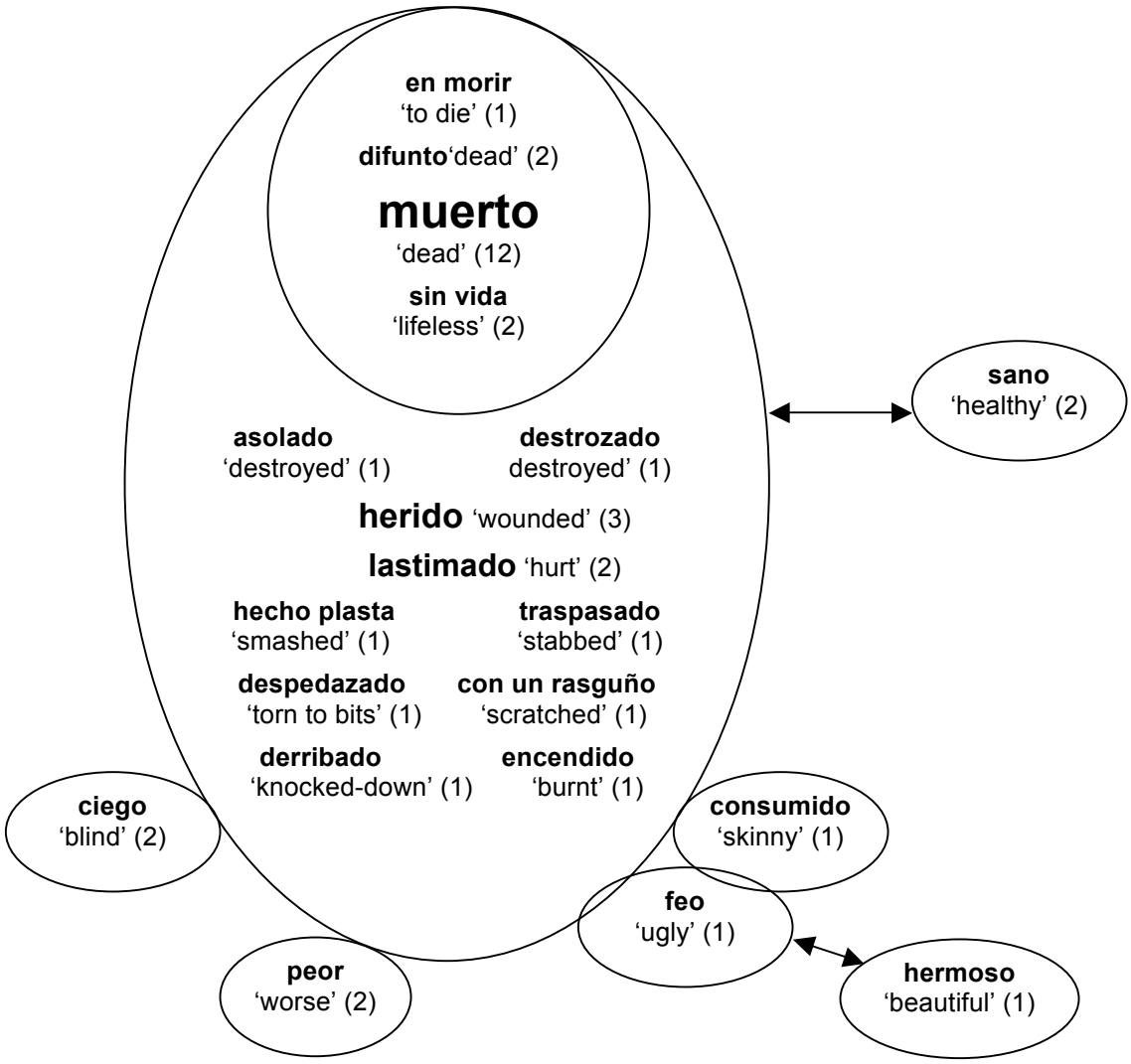


Figure 3, 1600's: Clusters centering on *muerto / herido*

Figure 3 is shown here in order to illustrate the methods of organization and the significance of the notation. This set of clusters comes from the *muerto* 'dead' clusters that were not examined in the analysis in Chapter 5. Instead, it is presented with the *muerto* clusters in Appendix 1.

On the list of adjectives for the 1600's, the adjective *muerto* 'dead' had a token frequency of 12 which was relatively high among adjective types in this century[7].

[7] Type and token frequency are discussed in more detail in Chapter 4.

Because of this, it was selected as a central member which is indicated textually in three ways. First, as in Bybee & Eddington (2006), it appears in a larger font as a way to iconically represent the fact that it would have a stronger mental representation than other less frequent exemplars. Second, it is placed in the center of the cluster that it occupies. Third, the number of tokens is given in parentheses so that the reader can see exactly how frequent it is in the corpus. Synonymous adjectives are located in the same immediate cluster as the central member because of the obvious semantic similarity. The adjective *difunto* is a direct synonym as evinced in the RAE (2005) definition: *Dicho de una persona: muerta (sin vida)* "Said of a person: dead (lifeless/without life)'.[8] Looking at Example (26) below, the prepositional phrase *en morir* 'to die' indicates that a change of state signaled by the verb *quedar* will occur that will leave the subject to die among his enemies. All of the exemplars in this cluster indicate that the change brought about is one that involves the finality of death.

(26) *Belisario: Y yo ruego al cielo santo, pues que vengarte deseas, que en los reinos africanos algún alarbe crüel, con alguna flecha o dardo, de Belisario la vida acabe, y así quedamos tú vengada y yo en morir entre mis fieros contrarios.*
'Belisario: And I beg holy heaven, since you desire to avenge yourself, that in the African kingdoms some savage, with some arrow or dart, ends Belisario's life, and this way we come out with you avenged, and me to die among my fierce opponents. (*El ejemplo mayor de la desdicha.* 17th C. Antonio Mira de Amescua. BVMC.)

In the oval dominated by the secondary central member *herido* 'wounded' there are adjectives that describe grievous bodily injuries that may result in death, but not with the certainty of *muerto*. A person could survive many of these states described, perhaps

[8] The fact that *sin vida* is mentioned in this definition is further evidence that this type is a closely related synonym, and that it refers to a state of being dead, as does the token extracted from Example (25) above in Section 3.2; the text reveals that the subject's enemies were killed in that example.

with the exception of *despedazado* ‘torn to bits’. The main thing that sets these verbs apart from *muerto* is the fact that they emphasize the manner of injury. Because of its higher token frequency, *herido* is considered the central member of these adjectives and they are arranged around it. However, the font of *herido* is not as large as *muerto* ‘dead’ because it appears with fewer tokens.

Because it doesn’t have a notably high relative token frequency, there is another factor that influences the proposition that *herido* is the central member of the cluster containing types denoting states of injury: generality. Central members tend to have a more general meaning; because of their generality they can appear in a variety of contexts and, therefore, tend to appear with higher token frequency. Furthermore, they are semantically related to a larger body of types than more specific types. In Figure 3, both *herido* ‘wounded’ and *lastimado* ‘hurt’ are types that display generality; they describe general situations that could apply to many situations. This is in contrast to more specific types, such as *despedazado* ‘torn to bits’ or *hecho plasta* ‘smashed’ that refer to specific ways of becoming injured. Even though *herido* and *lastimado* only have three and two tokens respectively, it is probably not a coincidence that their token frequency surpasses that of the other types in their immediate cluster; they are more general and, therefore, more useful.

The category, overall, contains adjective types that describe physical states that are typically detrimental or undesirable for the human organism. *Muerto* ‘dead’ and *herido* ‘wounded’ are related in a kind of scale in which *muerto* is the most extreme outcome of a physical change of state indicated by the construction *quedar(se)* + ADJ in this set of clusters. The adjective *muerto* and its other immediate category members are

enclosed in a cluster to indicate their similarity, while placed in a larger category that can include injury. Throughout the analysis, items in a circle are deemed to be either synonymous or very closely related in the context of use. Overlap between exemplar clusters is meant to show that the enclosed groups are related semantically even though they may not be synonyms. Seen this way, *muerto* and *herido* may not be synonyms, but they are related.

It must be remembered that these are theoretical category structures and that many of the details of how they are drawn up could be open to some debate. What cannot be debated is the fact that groups of semantically related adjectives of the nature shown in Figure 3 appear in the data over several consecutive centuries. As will be seen, this diachronic continuity is found in other sets of clusters as well. Because of this, the structure of the same set of clusters in surrounding centuries was a factor in determining proposed exemplar cluster organization.

Marginal members *ciego* 'blind', *peor* 'worse', *feo* 'ugly', and *consumido* 'skinny' are enclosed in their own bubbles and are shown touching the larger oval to indicate that they describe physical states that could be considered negative for the human organism but that they are not necessarily the result of an injury. They are located toward the bottom, instead of the top, to indicate that they are conceptually farther from the concept of death than the *herido* adjectives. These are marginal members that show a degree of relatedness to *herido* and even a thread of relatedness to *muerto* in that they are negative physical states. There were no other plausible sets of cluster to which they would relate and none has the significant token frequency that would justify its status as a central member of any other category. There is some continuity with following centuries

with the adjective *ciego*. It appears three times in the following century and its synonym *tuerto* 'blinded in one eye' appears in the 1800's.

Contextual analysis is extremely important in the placement of exemplar clusters as well. *Feo* 'ugly' and *consumido* 'skinny' may not seem as related to *herido* 'wounded' unless they are considered as adjectives participating in a change-of-state expression. One must consider that something happened to make this change of state occur. Seen as an outcome, it is easier to see how they both represent undesirable changes. As mentioned above, many of the tokens of the construction *quedar(se)* + ADJ have one single occurrence of the verb *quedar(se)* with more than one adjective. It is almost always the case that multiple adjectives appearing with a single occurrence of *quedar(se)* are related. Consider Example (27) in which the old lady, representing *vejez* 'old age', places a crown on her head which makes her appear even more ugly and skinny. Because of the context in which these two adjectives appear, they were placed in close proximity to indicate the fact that they were found in tandem with a single occurrence of the verb *quedar(se)*.

(27) *Por una estrecha puerta vio que entrava una vieja arrugada y carcomida, y que entre aquellas flores se assentava haziendo una guirnalda bien texida, con la qual la vejez se coronava quedando muy más fea y consumida.*
'Through a narrow door (he/she) saw that a wrinkled and eaten away old lady entered, and that among those flowers she settled making a well woven wreath, with which old age was crowned becoming much more ugly and skinny.' (*Genealogía de la toledana discreta*. 17th c. Eugenio Martínez. Lemir)

Two opposites, *sano* 'healthy' and *hermoso* 'beautiful', are also included in the clusters represented in Figure 3, and are indicated by double-headed arrows. This analysis takes the same stance on opposites as Bybee & Eddington (2006: 332) that "opposites share a number of features while having a negative value for one important feature". In

other words, opposites are semantically related. As indicated by the arrow, *hermoso* 'beautiful' is the opposite of the marginal adjective *feo* 'ugly', and *sano* 'healthy' is proposed to be opposite to all of the adjectives denoting a change of state in which the subject is hurt or killed.

There is one final point regarding the proposed category structure here and in the rest of the document. The notation used to denote possible clusters (i.e. the ovals and circles) represent theoretical category structures. If a different researcher were to modify the ovals, including some types and excluding others, it wouldn't actually affect the analysis much, as long as the adjectives maintained their general position. What cannot be refuted is that groups of semantically related adjectives were consistently used in the change-of-state construction *quedar(se)* + ADJ throughout several centuries.

Chapter 4: Overview of *quedar(se)* + ADJ through time

This section provides background information on the construction *quedar(se)* + ADJ regarding its emergence and endurance. This information will be pertinent throughout the detailed analysis of the different exemplar clusters of adjectives in the following chapter.

4.1 Overview

Table 3 is an overview of the proposed exemplar clusters. It shows the centuries in which the clusters appear and gives the section in which they are presented. For the purpose of organizing the material, the adjectives appearing in the construction *quedar(se)* + ADJ have been placed into three overarching categories: adjectives denoting social states, mental states, and physical states. These are not discrete categories and some of the adjective types are proposed to belong to more than one set of clusters. The analysis in Chapter 5 will cover the *solo* clusters, the *rico / pobre* clusters, the *libre* clusters, and the *confuso / suspenso* clusters. All others will be presented in Appendix 1[9]. The names of the clusters are based on the central members. If only one adjective is used in the cluster name (e.g. the *solo* clusters), it indicates that this particular adjective was the central member during the span of the data. If the cluster name involves two types (the *alegre / satisfecho* clusters), it indicates that at some point the central member shifted

[9] An in-depth analysis of the rest of the clusters would be beyond the scope of the present study.

Table 3. Overview of proposed exemplar clusters

	Social states				Mental state				Physical state	
	solo 'alone'	*libre* 'free'	*rico / pobre* 'rich / poor'	Other social states	*alegre / satisfecho* 'happy / satisfied'	*confuso / suspenso* 'confused / surprised'	*triste* 'sad'	transfer of info.	*muerto* 'dead'	body position
Section:	5.2.1	5.2.2	5.2.3	Appndx	5.3.1	Appndx	Appndx	Appndx	Appndx	Appndx
1200's	X	X	X	X	X	X	X		X	
1300's	X	X	X	X	X	X	X		X	
1400's	X	X	X	X	X	X	X		X	
1500's	X	X	X	X	X	X	X	X	X	
1600's	X	X	X	X	X	X	X	X	X	
1700's	X	X	X	X	X	X	X	X	X	X
1800's	X	X	X	X	X	X	X	X	X	X
1900's*	X				X	X	X	X		

* from Bybee & Eddington (2006)
X cluster of related adjectives appeared in the century indicated

One of the things that Table 3 demonstrates is that clusters of related adjectives have a tendency to show great endurance over time. At the same time, not all clusters have been present in the data in every time period; some disappeared and others emerged. Considering that each cluster consists of numerous adjective types, the emergence or disappearance of a set of clusters represents a significant change in the categories of adjectives used in *quedar(se)* + ADJ.

In some cases, this may have to do with emergent verbs of 'becoming'. For example, many of the adjectives having to do with possessing wealth, represented in Table 3 as the *pobre/rico* 'poor/rich' clusters, were not found in *quedar(se)* + ADJ in the 1700's or the 1800's. This reduction signaled the demise of these clusters; in the 1900's there was no evidence of their continuity in Bybee & Eddington (2006) or the *Corpus del español* (Davies 2002-). This could have to do with fact that many of these adjectives came to be used with a similar expression of 'becoming' such as *hacer(se)* + ADJ or *poner(se)* + ADJ (see Section 4.2.1.3, below). The *triste* clusters, and the *alegre /*

satisfecho clusters could have also been affected by the emergence of other verbs of 'becoming'. Even though Table 3 shows that these proposed clusters were present in the data in all centuries, many of their adjective types co-occurred with other change-of-state constructions. Bybee & Eddington (2006) found that many of the same adjectives that appear in the present data with *quedar(se)* + ADJ appeared in their 1900's data with other verbs of 'becoming' including types from the *pobre/rico* clusters, the *triste* clusters, and the *alegre / satisfecho* clusters. Adjectives denoting sadness and happiness (i.e. from the *triste* clusters, and the *alegre / satisfecho* clusters respectively) continued to appear in *quedar(se)* + ADJ into the 1800's and beyond, as documented here and by Bybee & Eddington. However, the central category member *triste* 'sad' came to be used with the change-of-state verb *poner(se)* by the 1800's (Balasch 2008). This is evidence that some of the changes in the adjective types appearing in exemplar clusters of *quedar(se)* + ADJ over the centuries also correspond with the emergence of other change-of-state verbs. This is discussed further in Section 4.2.1.3. In the earlier centuries (i.e. 1200's to 1400's), another verb of 'becoming' may have been a factor. Section 4.2.1.2 discusses the proposal that *quedar(se)* + ADJ may have inherited adjectives from an earlier expression of becoming, *fincar(se)* + ADJ.

This chapter also discusses the categorization of adjectives into exemplar clusters through time. There are two general tendencies in the exemplar clusters presented in this study. Some, such as the *solo* clusters, show endurance throughout the centuries with very little modification to the overall structure of the clusters and a persistent presence of the central member. Others, such as the *alegre / satisfecho* clusters, mutated significantly through time, showed shifts in central members, and had significant changes in type

frequency. Furthermore, this chapter presents evidence that there is a correlation between token frequency and type frequency. As high frequency central members emerge in the data, the category expands with the appearance of new types in the data.

4.2 Commentary regarding the exemplar model in this diachronic study

Because this study employs the exemplar model in order to account for categorization, the clusters illustrated here by theoretical 'bubbles' are not meant to represent discrete categories in which an adjective type is locked into only one set of clusters or just one category. Furthermore, these exemplar clusters represent fuzzy, flexible boundaries that change over the years by expanding or contracting. Some adjective types are proposed to belong to more than one set of clusters and some types have been found to co-occur with other verbs of 'becoming'.

These proposed categories of exemplar clusters are based on speaker experience with language and, as a result, would be modified slightly each time a speaker experiences a token of *quedar(se)* + ADJ in the input or output by reinforcing existing representation or by expanding into new territory. The categories that are formed are used not only to interpret incoming items but affect the probability that one form will be produced over another

The data used for this study come from a written corpus and writing is a form of production, or linguistic output. This collective output drawn from the numerous writers in the corpus, according to the exemplar model, is the result of these writers' experiences with language and the categories that they have formed as a result. Given that each experience slightly modifies the category on an individual level, it is important to put these linguistic experiences into context. First, as many researchers have observed (e.g.

Zuraw 2003) in treating linguistic representation and probability, language change happens through speaker interactions. These modifications to the grammar are based on previous usage.

This leads to the second point: under ordinary life circumstances these previous experiences with language are drawn by an individual from a group of speakers, a speech community. Furthermore, there is a dynamic between individuals and the group in which modifications happen at the individual level and are spread to the speech community. With this in mind, novel usage of a particular adjective in *quedar(se)* + ADJ will typically reflect an individual's innovation based on analogical extension to previously experienced adjective types. If a novel form is spread throughout the group, it may undergo a rise in token frequency and endure over several centuries.

4.2.1 Overview of *quedar(se)* + ADJ types, tokens, and frequency through time

Since the bulk of the following chapter focuses on presenting individual exemplar clusters as they progress through the centuries, it is necessary to place these into a larger scheme concerning some of the overall trends of the construction *quedar(se)* + ADJ. In the data, there are fluctuations in the overall frequency of *quedar(se)* + ADJ and in the ratio of tokens per adjective type. These figures are relevant in the upcoming analysis of the different exemplar clusters and there are many references made to this section. In order to recapitulate some of the information presented in Table 2, above, Table 4 provides the number of tokens found in the data, the number of adjective types, the standardized frequency of *quedar(se)* + ADJ calculated as the number of tokens per 10,000 words, the token/type ratio, and the number of words in the corpus per century. The figures for the 1900's were calculated using data from Bybee & Eddington (2006).

They consulted a spoken corpus of approximately 1.1 million words of Peninsular Spanish and a written corpus of 990,000 words from fifteen novels written by authors from various regions. In order to be consistent with the written mode of my own corpus of written Spanish, the calculations from Bybee & Eddington (2006) presented here come from their corpus of written Spanish only.

Table 4. Overview of *quedar(se)* + ADJ types, tokens, and frequency through time

	Tokens	Types	Frequency: # per 10,000	Token/Type ratio	# words in corpus
1200's	43	29	0.42	1.48	1,031,755
1300's	76	40	0.96	1.9	789,426
1400's	196	121	1.18	1.62	1,658,165
1500's	271	153	4.85	1.77	558,475
1600's	280	171	6.00	1.63	463,529
1700's	253	135	5.82	1.87	435,055
1800's	255	141	2.64	1.80	967,252
1900's*	181	54	1.83	3.35	990,000

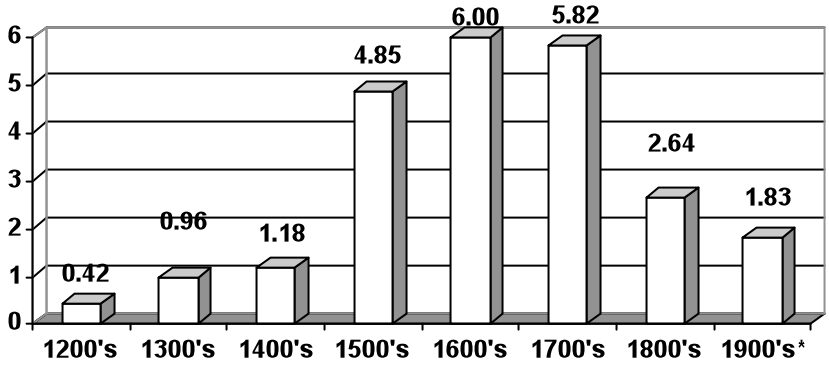
* From Bybee & Eddington (2006)

For more information on the low number of tokens in the span from the 1200's to the 1400's, see Chapter 3:, Sections 3.1 and 3.2. In calculating the number of types, all morphological manifestations of a particular adjective were considered to belong to one type (e.g. feminine singular *sola,* and masculine plural *solos* would be tokens of the adjective type *solo* 'alone'). As is discussed below in Chapter 5:, Section 5.1.1, it facilitates the discussion to use the generalizations *quedar(se) sin* + NOUN 'to be left without + NOUN' and *quedar(se) con* + NOUN 'to be left with + NOUN' to refer to the myriad of types that appear in the construction *quedar(se)* + ADJ with the prepositions *sin* and *con* (e.g. *sin padre / ojos / honra* 'without a father / eyes / honor'). However, in calculating type frequency, each different combination of *quedar(se) sin* + NOUN and *quedar(se) con* + NOUN that appears in the data is considered one type. Even though *sin*

padre ‘w/o father’ and *sin maestro* ‘w/o teacher’ both appear in the 1700’s in a cluster named *sin* + HUMAN, each is considered to be a different type (see Table 11, p. 103).

The Frequency column shows the overall standardized frequency of *quedar(se)* + ADJ for each century determined by the number of occurrences per 10,000 words. As stated in Chapter 3, texts were loaded into the corpus in their entirety and the calculation of overall frequency for *quedar(se)* + ADJ is based on the total number of words in all of the texts used in each century. Chart 1 gives a visual representation for the overall standardized frequency of *quedar(se)* + ADJ in the eight centuries with data from the 1900’s coming from Bybee & Eddington (2006).

Chart 1. Overall frequency of *quedar(se)* + ADJ in 8 centuries; occurrences per 10,000 words.



* Calculated from data from Bybee & Eddington (2006)

The overall standardized frequency of the *quedar(se)* + ADJ in the time span studies shows a sort of bell curve. It has a relatively low frequency in the 1200’s (0.43 tokens per 10k words). The frequency increases significantly by the 1500’s (4.85 per 10k) and peaks in the 1600’s (6.00 per 10k). By the 1900’s the overall frequency (1.83 per 10k) has dropped down again. In the earlier centuries, this could be due to the fact that a similar expression of ‘becoming’ *fincar(se)* + ADJ was prominent in usage, as is

discussed in Section 4.2.1.2. Similarly, the decline in the overall frequency could be attributable to the fact that many adjectives used in *quedar(se)* + ADJ had come to co-occur with other emergent expressions of 'becoming' as discussed below in Section 4.2.1.3.

Returning to Table 4, the column with the token per type ratio provides relevant information to the analysis of frequency in *quedar(se)* + ADJ. This ratio gives insight into novel usage versus established usage. If the token per type ratio is lower (i.e. fewer tokens per type), it is an indication that there are more novel forms and that there are fewer types with a high token frequency. If the ratio is higher (more tokens per type), it indicates that there is a tendency for tokens to be concentrated in fewer types. The 1200's, for example, has the lowest token per type ratio (1.48) because types are spread across many low frequency types and there are few types with a high token frequency; the average number of tokens per type was 1.48, and 19 out of a total of 29 types (65.5%) had only one token in the data in this century. In the following century, the 1300's, there was an increase in overall frequency of *quedar(se)* + ADJ (from 0.42 per 10k words in the 1200's to 0.96 in the 1300's) and an increase in the token per type ratio to 1.9 (i.e. an average of nearly two tokens per type). Usage is more concentrated in this century; only 24 out of 76 types (31.6%) had just one token. All other types had at least two tokens showing some evidence of conventionalization.

4.2.1.1 Prefabs and conventionalized instances of constructions (CIC's)

One of the claims that this dissertation advances is that prefabs have longevity. This is logical considering that prefabs are ubiquitous in language use. Compared to less formulaic sequences, Erman & Warren (2000: 50) propose that prefabs constitute about

55% all written and spoken discourse[10]. This section proposes one way of arriving at a prefab threshold using relative frequency which will be applied only to two prominent types: *quedar(se) solo* 'to be left alone' and *quedar(se) libre* 'to be set free'. Furthermore, this section proposes a threshold for identifying conventionalized instances of constructions (CIC's)[11] that will be applied more exhaustively.

The aim of developing a CIC threshold is to have a method that is readily applicable to the set of data at hand in order to show the prevalence of formulaic combinations without going beyond the scope of the present study. The method for identifying CIC's makes use of relative frequency of combinations of *quedar(se)* + ADJ in my own data and is used in order to make general statements about the correlations between frequency and formulaicity. As Bybee & Torres Cacoullos (2008) propose, instead of a discrete partition between fixed expressions and productive formations, there is a continuum. CIC is a general and inclusive term; many of these could be routinized enough to be prefabs. The main claim about CIC's however, is that they are specific combinations of *quedar(se)* + ADJ that show evidence of formulaicity and are on the more fixed end of the continuum. They are a conventionalized manner of expressing a certain change of state.

Only relative frequency is used to operationally identify CIC's. The token per type ratio provides a convenient figure for proposing a CIC threshold since it represents the average number of tokens per type. If a certain type has twice the average number of tokens for the century in which it appears, I propose that this is evidence of

[10] Erman & Warren (2000) also acknowledge that it is difficult to categorically identify prefabs and that their measure may be conservative. Therefore, prefabs may account for more than 55% of discourse.
[11] This term "conventionalized instances of constructions" comes from Bybee & Torres Cacoullos (2008: 3), though they did not use it as a specific term regarding a defined group of types.

conventionalization compared to types with average, or lower than average, token frequency. Therefore, the proposed formula for determining the threshold is: (Token/Type ratio) x 2, rounded to the nearest whole number. This method takes into account patterns of distribution in the data for each century; some centuries have more scattered adjective usage among many novel types and some show a tendency for types to be more concentrated in high frequency central members, as discussed in the previous section. Table 5 gives the token per type ratio, the proposed CIC threshold, the number of types that have a token frequency at, or above, the threshold, and a list of these types.

Table 5. Overview of *quedar(se)* + ADJ CIC threshold and types

	Token/Type ratio	CIC Thrshld	CIC Types	Types
1200's	1.48	3	3	*por heredero* 'as lord' (3), *sin señor* 'w/o lord' (3), *solo* 'alone' (3)
1300's	1.9	4	4	*confuso* 'confused' (7), *por señor* 'as lord' (6), *señor* 'lord' (4), *solo* 'alone'(4)
1400's	1.62	3	16	*alegre* 'happy' (4), *concorde* 'in agreement' (5), *herido* 'wounded' (3), *muerto* 'dead' (4), *obligado* 'obligated' (6), *por gobernador* 'as governor' (3), *por rey* 'as king' (5), *por señor* 'as lord' (4), *pobre* 'poor' (3), *preñada* 'pregnant' (4), *sano* 'healthy' (4), *sucesor* 'successor' (4), *solo* 'alone' (8), *triste* 'sad' (8), *vencedor* 'victor' (3), *viudo* 'widowed' (6),
1500's	1.77	4	14	*burlado* 'mocked' (6), *ciego* 'blind' (4), *confuso* 'confused' (6), *contento* 'happy' (7), *espantado* 'scared' (5), *libre* 'free' (11), *muerto* 'dead' (12), *obligado* 'obliged' (6), *preso* 'imprisoned' (4), *quejoso* 'whiny' (4), *sano* 'healty' (9), *satisfecho* 'satisfied' (6), *sin él* 'w/o him' (3), *solo* 'alone' (13)
1600's	1.63	3	24	*absorto* 'astonished' (3), *alegre* 'happy' (3), *atónito* 'astounded' (6), *confuso* 'confused' (3), *contento* 'happy' (7), *corrido* 'embarrassed' (6), *desengañado* 'disappointed' (3), *disculpado* 'forgiven'(4), *embelesado* 'fascinated' (4), *fuera de sí* 'outside himself' (3), *herido* 'wounded' (4), *libre* 'free' (8), *mudo* 'mute' (3), *muerto* 'dead'(12), *perdido* 'lost' (3), *rendido* 'exhausted' (4), *rico* 'rich' (4), *satisfecho* 'satisfied' (9), *seguro* 'safe' (4), *solo* 'alone' (10), *triste* 'sad' (5), *ufano* 'cheerful' (4), *vencido* 'defeated' (4), *vengado* 'avenged' (4)
1700's	1.87	4	14	*admirado* 'amazed' (8), *atónito* 'astounded' (8), *aturdido* 'bewildered' (4), *confuso* 'confused' (4), *desmayado* 'fainted' (6), *dormido* 'asleep' (10), *inmóvil* 'immobile' (6), *lastimado* 'hurt' (4), *pagado* 'satisfied' (4), *pasmado* 'stunned' (10), *satisfecho* 'satisfied' (7), *solo* 'alone' (5), *sorprendido* 'surprised' (6), *suspenso* 'surprised' (12)
1800's	1.80	4	10	*convencido* 'convinced' (4), *dormido* 'asleep' (8), *enterado* 'informed' (4), *inmóvil* 'immobile' (4), *libre* 'free' (6), *huérfano* 'orphaned' (7), *mudo* 'mute' (6), *satisfecho* 'satisfied' (6), *solo* 'alone' (29), *suspenso* 'surprised' (7)

The proposed CIC threshold is useful in providing an overview of how many of the types and tokens in the data are members of conventionalized combinations. Table 6 shows how many types there are in the data, how many of these types are CIC's, the total

number of tokens in the data, the number of tokens that are from CIC types, and the overall frequency of *quedar(se)* + ADJ per 10k words. For example, 14 (10.5%) of all types found in the 1700's had a token frequency at or above the CIC threshold, and these account for 94 (37%) of all of the tokens in the data for this century.

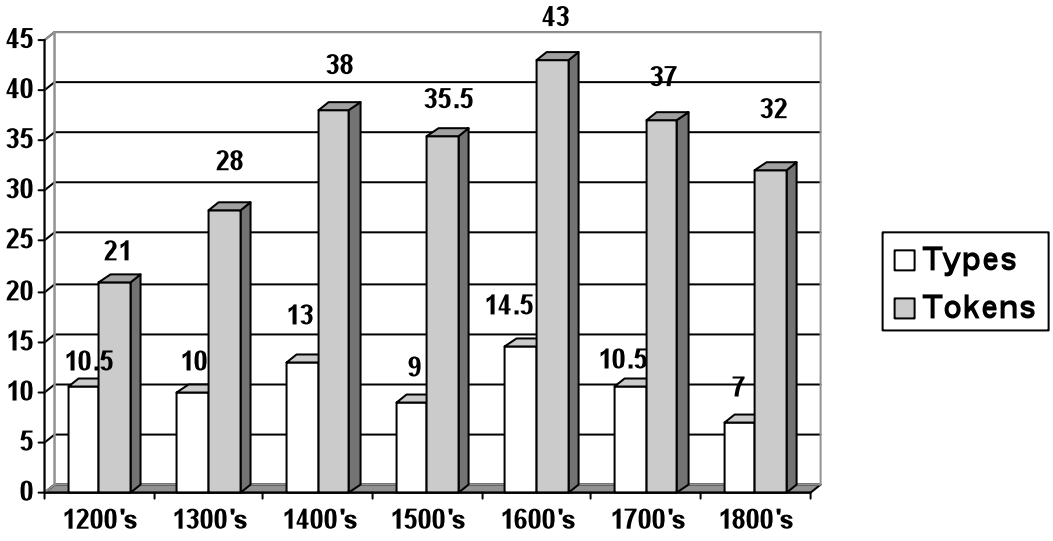
Table 6. Overview of *quedar(se)* + ADJ types and tokens above the CIC threshold

	N total types	N CIC types (%)	N total tokens	N CIC tokens (%)	Frequency: # per 10,000
1200's	29	3 (10.5%)	43	9 (21%)	0.42
1300's	40	4 (10%)	76	21 (28%)	0.96
1400's	121	16 (13%)	196	74 (38%)	1.18
1500's	153	14 (9%)	271	96 (35.5%)	4.85
1600's	171	24 (14.5%)	280	119 (43%)	6.00
1700's	135	14 (10.5%)	253	94 (37%)	5.82
1800's	141	10 (7%)	255	81 (32%)	2.64

Table 6 shows that there is a correlation between the number of CIC's identified in the data and the overall frequency of *quedar(se)* + ADJ. The 1200's is the century with the lowest standardized frequency of *quedar(se)* + ADJ (0.42 per 10k words) and has the lowest percentage of CIC's with only 21% of all tokens occurring above the threshold. The century with the highest overall frequency, the 1600's (6.00 occurrences per 10k words), has the highest percentage of tokens (43%) and the highest percentage of types (14.5%) that are CIC's. It makes sense that CIC's would occur at the highest rate in the data in the century in which *quedar(se)* + ADJ is being used the most; by occurring at a higher overall frequency there is more of an opportunity for conventionalization and there is a greater number of frequently occurring combinations in usage. As the overall frequency of *quedar(se)* + ADJ diminishes in the 1700's and 1800's, the percentage of tokens (37% and 32% respectively) and types (10.5% and 7% respectively) accounted for by CIC's diminishes accordingly. Chart 2 provides a more visual account of the

corresponding rise in the percentage of tokens and types in the data accounted for by CIC's.

Chart 2. Percentage of adjective types and tokens accounted for by CIC's.



On the other hand, the method for identifying prefabs uses a much larger set of data, the *Corpus del español* (Davies 2002-), and takes relative frequency as well as intervening syntactic material[12] into account. The reason for developing a method of identifying prefabs, given that much of the discussion in the following chapter will deal with CIC's, is twofold. First, this measure is used in order to advance the claim that prefabs have longevity. Second, it provides a methodology for arriving at an operational definition of prefabs that can be applied to constructions such as *quedar(se)* + ADJ in future research. Since it makes use of an outside set of data, it is only applied to two types

[12] In developing a prefab threshold, the search for instances of *quedar(se) solo* and *qudar(se) libre* on Davies' (2002-) data was for combinations with no intervening words. By controlling this syntactic factor, it contributes to the notion that the verb and the adjective are stored together as a unit and may be an indication of unithood (Bybee & Torres Cacoullos 2009).

found in these data and is considered to be more of a methodological step instead of an exhaustive identification of all prefabs found in the data.

Bybee & Eddington (2006) and Bybee (2006), in discussing verb + adjective change-of-state constructions, proposed that types with high token frequency were prefabs. Although no specific threshold was proposed, nor was it intended as an exhaustive list, Bybee (2006) listed the following verb + adjective combinations as being prefabs based on their conspicuously high token frequency relative to other combinations: *ponerse nervioso* 'to get nervous' (N = 17), *quedarse solo* 'to end up alone' (N = 28), and *quedarse inmóvil* 'to be immobile'(N = 17). In referring to these combinations, Bybee & Torres Cacoullos (2009: 191) state that these combinations are prefabs "in that they represent the normal, conventionalized way of expressing commonly referred-to changes of state".

In order to identify prefabs, Bybee & Torres Cacoullos (2009: 204) consider relative frequency instead of overall frequency, and arrive at the following operational prefab threshold for "auxiliary" + gerund combinations in Spanish: "combinations making up 2% or more of the corresponding "auxiliary" data and 50% or more of the corresponding gerund data". In their ensuing example, the combination *estar hablando* 'to be talking' would be a prefab because it accounts for 5% (32/672) of all occurrences of the auxiliary verb *estar* and 71% (32/45) of all occurrences of the gerund *hablando* in their data.

Because of the breadth and scope of the data, this study uses the *Corpus del español* (Davies 2002-) in order to arrive at a proposed prefab threshold. Adapting some of the rationale in Torres Cacoullos' (2001: 59) study of auxiliary-plus-gerund sequences

to the present data, it is important to measure the adjective not only as part of these combinations, but against overall adjective usage as well. Also, increased construction frequency makes it more likely that an adjective is tied to a certain change-of-state verb. I operationally define prefabs as combinations of *quedar(se)* + ADJ in which (a) the adjective accounts for at least 1% of all occurrences of that adjective in Davies' (2002-) data, (b) the combination accounts for at least 4% of all occurrences of *quedar(se)* + ADJ in the same data, and(c) there is no intervening material between the verb and the adjective.

These figures for arriving at an operational definition differ from Bybee & Torres Cacoullos' (2009) because the patterns of distribution for the two constructions (*quedar(se)* + ADJ and "auxiliary" + gerund) are different. As will be seen with *quedar(se) solo* 'to be left alone' (Section 5.2.1.8, Table 12) and *quedar(se) libre* 'to be set free' (Section 5.2.2.8, Table 16), in no century do either of these combinations comprise more than 5% of all occurrences of the adjective nor more than 14% of all occurrences of *quedar(se)* + ADJ in Davies' (2002-) data. Furthermore, many of the adjectives that occur in *quedar(se)* + ADJ are more frequent than gerunds in usage. For example, in the 1500's the *Corpus del español* (Davies 2002-) shows that the gerund *esperando* 'waiting' occurred at a rate of 60.6 times per million words; the combination of this gerund with *estar* 'to be' was fairly robust comprising 7% of the "auxiliary" data and 89% of the gerund data. In the same century, *solo* 'alone' occurred at a rate of 365.1 times per million words and *libre* 'free' at a rate of 229.1 per million. In light of this, the figures I propose represent reasonable cutoffs based on the distribution of the data that give a maximal number of prefabs while excluding less conventionalized combinations.

This operational definition will be applied to the central members *solo* 'alone' and *libre* 'free' and will be presented in Sections 5.2.1.8, and 5.2.2.8 respectively.

4.2.1.2 The rise of *quedar* and the demise of *fincar*

There is a factor that may play a part in the low frequency of *quedar(se)* + ADJ in the earlier centuries: the concurrent, but waning, usage of the change-of-state construction *fincar(se)* + ADJ. This section demonstrates how the changing frequency of these two very similar verbs may be attributed to the fact that *quedar* was coming to replace *fincar* in a variety of uses. Further consideration of these two verbs may shed some light on this issue.

In order to investigate the semantic relationship between *quedar* and *fincar* in a diachronic setting, I used the *Corpus del español* (Davies 2002) as a database for this analysis gathering data from alternating centuries beginning with the 1200's (i.e. 1200's, 1400's, 1600's, and 1800's). Because it was possible that some meanings of the verbs were associated with certain verb tenses, I obtained a representative sample reflecting the ratio of each verbal tense to its total occurrences in the data of approximately 200 verbs for each century studied. For example, if, in a given century, 42% of all occurrences of *quedar* were in the preterit and 35% in the imperfect and so on, I collected 84 tokens (i.e. 42% of 200) in the preterit, 70 (i.e. 35% of 200) tokens in the imperfect, and so on. However, because of the diminishing frequency of the verb *fincar* (see Table 8 below), I was able to collect and analyze all tokens in the last two centuries analyzed, the 1600's (N=58) and 1800's (N=14).

The purpose was to see what happened to these verbs through time in regards to the semantic domains occupied by the verbs and to see how this changed diachronically.

Pountain (2001: 56) observes that "*fincar* was common in Old Castilian in the meaning of 'to remain', and was one of a number of semantically weakened verbs denoting little more than 'being (in a place)'." The Latin root of *fincar* is [FĪGĬCO] meaning 'to fix' or 'to nail.' He goes on to list *quedar(se)* as another of these semantically weakened verbs and lists its Latin root as [QUIETO] 'to quieten' (Pountain 2001). Both of these verbs expressed more than one meaning while, at the same time, maintaining a certain degree of their Latin meaning. Perhaps the semantic weakening mentioned by Pountain was based on a reaction to the fact that these verbs are very generalized in their usage.

Examples (28) and (29) below demonstrate how the two verbs could mean 'to remain in a place' (with the locative PP, *en* + NP).

(28) *Finco Nabuchodonosor en manip muchos dias.*
'Nabuchodonosor stayed in Manip many days.' (General estoria IV. 13th c. Alfonso X. Davies 2002)

(29) *Los ricos hombres que hauian quedado en la tierra supieron como los de bellinas estauan en muy gran fatiga.*
'The rich men who had remained in the land found out how those from Bellinas were very fatigued' (Gran conquista de Ultramar. 13th c. Anónimo. Davies 2002)

Examples (30) and (31) are tokens that display the usage of the two verbs with an adjective and a human subject to express a change of state. These are the change-of-state constructions *fincar(se)* + ADJ, and *quedar(se)* + ADJ.

(30) *Vendian a sus vezindades: en guisa que fincaron muy ricos de alli adelante.*
'They sold their surrounding land: this is how they became rich from then on.' (Crónica del Cid, Anónimo 15th c.)

(31) *E en aquella primera noche delas bodas que el conde & la condessa durmieron queda ella preñada.*

'And in that first night of the weddings that the count and the countess slept (together) she becomes pregnant.' (Gran conquista de Ultramar, Anónimo, 13th c.)

As seen in the following examples, (32) and (33), both verbs could also mean 'to stop', much as in the usage of the modern *dejar*. It is possible that these examples are representative of a more specific construction *fincar / quedar de ir* 'to stop going'.

(32) *Maguer que este Rey don Alffonsso era muy ninno. por tod esso. non finco de yr el muy derrezio contra los aragoneses.*
'Even though this King don Alfonso was very child(like). For this. He didn't stop going very strongly against the Aragonites.' (Estoria de España II, Alfonso X, 13th c.)

(33) *Et non quedo fasta que alli llego. & fizo aquello. Et nin querie quedar aun de yr adelant mas; fasta que pleytearon con el; el Rey de Toledo. & el Rey de Cordoua.*
'And he didn't stop until he arrived there. And did that. And he still didn't want to stop going ahead more; until they made a pact with him; the King of Toledo. And the King of Cordova.' (Estoria de España II, Alfonso X, 13th c.)

In the following examples, *quedar* is used both intransitively, in (34), and transitively, in (35), to mean 'to quieten,' or 'to calm / pacify' echoing its Latin predecessor [QUIETO]. This meaning is the one conveyed by modern *quietar*, a verb that does not presently convey any other meanings (DRAE 2006).

(34) *Quando queda la fiebre el enfermo finca suelto & manso.*
'When the fever calms, the sickness becomes loosened and docile.' (Tratado de las fiebres, Isaac Israeli, 15th c.)

(35) *Non la pueden quedar nin amansar.*
'They can neither calm nor tame her.' (General estoria V, Alfonso X 15th c.)

In example (36) *fincar* reflects its Latin root [FĪGĬCO] meaning 'to fix something to something' or 'to nail.' However, this example may reflect a possible extension of

these two meanings in early centuries as it could also mean 'to place.' One of the ways that this usage can be conceptualized in English is by the verb 'to stick' which can be used in a similar manner (i.e. 'Don't stick that bean in your ear,' 'I'm sticking this poster on the wall,' or 'stick that bowl on the table').

(36) *En algunos lugares es nesçesario fincar alguna cosa en cosa que non es mucho dura.*
'In some places it is necessary to place something in something that is not very hard.' (Cirugía mayor, Lanfranco de Milano, 15th c.)

Furthermore, *fincar* was used in two prominent fixed expressions, or prefabs. One was 'to set up a tent' (*fincar una tienda*) as in example (37) below. The other fixed expression was 'to kneel' (*fincar la rodilla*) as in example (38).

(37) *E el rey bucar mando fincar las tiendas enderredor de valencia.*
'And King Bucar gave the order to set up the tents around Valencia.' (Crónica del Cid, Anónimo 15th c.)

(38) *& oliueros finco la rodilla en el suelo & rescibio el gracioso beso.*
'And Oliveros kneeled upon the floor and received the gracious kiss.' (Oliveros de Castilla, Anónimo, 15th c.)

One of the key observations is that *quedar* and *fincar* changed significantly over the years in regards to the meanings that they conveyed and in their overall frequency. Table 7 shows the distribution of the verbs in the semantic domains that they occupied in each of the four centuries studied.

Table 7. Meaning and distribution in four centuries for *quedar* and *fincar*.

	1200's		1400's		1600's		1800's	
	quedar	*fincar*	*quedar*	*fincar*	*quedar*	*fincar*	*quedar*	*fincar*
To remain	131 (64%)	115 (58%)	147 (73%)	96 (50%)	107 (55%)	27 (62%)	108 (52%)	8 (62%)
To become	9 (5%)	22 (11%)	42 (21%)	27 (14%)	86 (44.5%)	14 (32%)	99 (48%)	2 (15%)
To stop	45 (22%)	3 (1.5%)	3 (1.5%)	-	1 (0.5%)	1 (2%)	-	-
To calm down	19 (9%)	-	9 (4.5%)	-	-	-	-	-
To stick	-	25 (12.5%)	-	20 (10%)	-	1 (2%)	-	3 (23%)
To kneel	-	17 (9%)	-	42 (22%)	-	1 (2%)	-	-
To set-up a tent	-	15 (8%)	-	7 (4%)	-	-	-	-
Total	204	197	201	192	194	44	207	13

In the 1200's, 5% (9/204) of all tokens of the verb *quedar* are tokens of the change-of-state construction, *quedar(se)* + ADJ ('to become' in Table 7), in the *Corpus del español* (Davies 2002-). However, 11% (22/197) of all tokens of *fincar* mean 'to become'. By the 1400's, there is evidence that the construction *quedar(se)* + ADJ is gaining in frequency relative to other uses of the verb *quedar* as 21% of all tokens of *quedar* in this century convey a change of state (four times as many as that in the 1200s). In the same century 14% of all tokens of *fincar* are used to express 'becoming', a rise of slightly over one quarter from 11% in the 1200s. It's also interesting that the 'remain' use rises for *quedar* but drops for *fincar*.

Evidence for the expansion of *quedar(se)* + ADJ relative to other uses of *quedar* is seen in the 1600's and the 1800's as it respectively represents 44.5% and 48% of the total tokens of *quedar* in the data from these centuries. In these two centuries, not only is *fincar* used at a lower rate than *quedar* to express a change of state (32% and 15% respectively), but the overall number of tokens of *fincar* is diminishing (44 and 13 in the

data, respectively). The fact that I was not able to obtain the target number of tokens is due to the loss in overall frequency of the verb *fincar*, as documented in Table 8 below.

Table 8. Overall frequency of the verbs *fincar*, *hincar*, and *quedar* from the 1200's to the 1800's in the *Corpus del español* (Davies 2002).

	1200's		1400's		1600's		1800's	
	# per 10,000	N	# per 10,000	N	# per 10,000	N	# per 10,000	N
Fincar	5.91	4087	0.95	810	0.05	58	0.007	14
Quedar	1.17	808	3.77	3209	10.62	13,545	7.96	16,298
Total # words in *Corpus*	6,905,000		8,515,000		12,746,000		20,465,000	

The verb *fincar* appears in the data less frequently as time passes. It starts out with an overall frequency of 5.91 occurrences per 10k words in the 1200's and ends up with a frequency of 0.007 occurrences per 10k words in the 1800's. During the same time span, *quedar* goes from having an overall frequency of 1.17 tokens per 10k words in the 1200's and increases to 10.62 per 10k words in the 1600's at which point it tapers off slightly ending up at 7.69 tokens per 10k words in the 1800's. This 'bell curve' of the overall frequency of the verb *quedar* in Table 8 resembles the frequency curve for the change-of-state construction *quedar(se)* + ADJ from Table 4 and Chart 1 above. In other words, it is not just *quedar(se)* + ADJ that is diminishing in frequency, but the overall usages represented by the verb *quedar*.

Taken together, Table 7 and Table 8 show that as *quedar* gained in overall frequency, a higher percentage of the total tokens was accounted for by the expression of 'becoming' *quedar(se)* + ADJ. At the same time, the verb *fincar* diminished in frequency. This raises other issues. First, the rise in the overall frequency of *quedar(se)* + ADJ must be related to the gain in the overall frequency of the verb *quedar*, and vice versa. The

gain in frequency of the verb *quedar* as a whole (as seen in Table 8) seems to correlate with gains in the frequency of *quedar(se)* + ADJ (as seen in Table 4). Furthermore, the decline in overall frequency of *quedar(se)* + ADJ from the 1600's to the 1900's seems to correlate with the tapering off of the overall frequency of the verb *quedar*. This points to a scenario that is similar to the expansion and contraction of adjective categories used in *quedar(se)* + ADJ; the rise in token frequency of a verb is accompanied by a rise in the types of contexts in which it may be used and the decline in token frequency is accompanied by a reduction in contexts of usage. Also, this points to a broader process of language change in which a form, or a combination, undergoes fluctuations in overall frequency that are related to expansion of usage and contraction of usage. A form may be in the process of becoming more frequent, or it may be in the process of disappearing in usage.

Second, in accounting for the fact that there was a gain in the overall frequency of *quedar(se)* + ADJ from the 1200's to the 1600's and a loss from the 1600's to the 1900's, it must be considered how other expressions of 'becoming' have related to these fluctuations. Specifically, the adjectives used in other change-of-state constructions must be considered. In looking at the 1200's and the 1400's, the two centuries before the conspicuous decline of the overall frequency of *fincar* to almost negligent levels, we see that, when used as an expression of 'becoming', there are many adjectives that co-occur with *quedar(se)* + ADJ. Table 9 lists all of the adjectives used in the change-of-state construction *fincar(se)* + ADJ with an animate subject that also appear in *quedar(se)* + ADJ.

Table 9. Adjectives co-occurring with *fincar(se)* + ADJ and *quedar(se)* + ADJ: 1200's & 1400's.

asegurado 'assured'	*por rey* 'as king'
assolado 'destroyed'	*rico* 'rich'
burlado 'mocked'	*seguro* 'sure'
contento 'happy'	*señores* 'lords'
desamparado 'separated, unprotected'	*solo* 'alone'
deshonrado 'dishonored'	*triste* 'sad'
libre 'free'	*vencido* 'defeated'
muerto 'dead'	*vivo* 'alive'
pagado 'payed/satisfied'	The general structure: *sin* 'without' + NOUN
pobre 'poor'	

There were also other adjectives that occurred in *fincar(se)* + ADJ that, though are not identical, share semantic similarity with adjectives appearing in *quedar(se)* + ADJ in this study. The following adjectives appeared in *fincar(se)* + ADJ but could have plausibly appeared in *quedar(se)* + ADJ based on similarity to types found in the data: *desheredado* 'disinherited', *dominado* 'dominated', *enflaquecido* 'skinny', *espavorido* 'terrified', *helado* 'frozen (both physically and metaphorically)', and *sañudo* 'irate'. In fact, almost all of the adjectives used with a human subject in *fincar(se)* + ADJ bear a semantic similarity to those in *quedar(se)* + ADJ. Coupled with the fact that *fincar* and *quedar* share synonymous uses, including the dichotomy of meaning both 'to remain' and 'to become' (i.e. show similar patterns of polysemy), it is possible to propose a scenario in which one synonym became more frequent than the other, as will be demonstrated to happen with many adjective types in the upcoming analysis. It is likely that as *fincar* diminished in frequency many of its functions were attributed to *quedar* as it rose in frequency. This is supported by the observation that many of the same adjectives appearing in *fincar(se)* + ADJ either co-occurred with, or were later used in *quedar(se)* + ADJ. Perhaps, as *fincar(se)* appears to have been replaced by *quedar(se)*, another verb

with a similar meaning, such as *permanecer* 'to stay / remain', may come to replace *quedar(se)* as a verb of 'becoming' and in other contexts.

4.2.1.3 The endurance of *quedar(se)* + ADJ and the emergence of other expressions of 'becoming'

In the more recent centuries studied, there is evidence of interaction between the frequency of *quedar(se)* + ADJ and other emergent expressions of 'becoming'. Referring back to both Table 4 and Chart 1, the overall frequency of *quedar(se)* + ADJ declines from the 1600's to the 1900's. It is likely that this is partially due to the expansion of other verbs of 'becoming' such as *ponerse, hacerse,* and *volverse*. These were the other verbs of 'becoming' studied in Bybee & Eddington (2006). To my knowledge, there are no diachronic studies of *hacerse* and *volverse* to which references can be made, rendering them beyond the scope of this discussion. However, Balasch (2008) studied *poner(se)* + ADJ as a change-of-state expression in the 1600's and the 1800's. Table 10 gives the adjectives that co-occur with *poner(se)* + ADJ in Balasch's study and *quedar(se)* + ADJ in the present study.

Table 10. Adjectives appearing with both *poner(se)* + ADJ and *quedar(se)* + ADJ in the 1600's and 1800's.

contento 'happy'*	*pálido* 'pale'
delicado 'delicate'	*peor* 'worse'
encendido 'lit up (in face, with love)'	*serio* 'serious'*
enfermo 'sick'	*trémulo* 'trembling'
fuera de sí 'beside him/herself'	*triste* 'sad'*
furioso 'furious'	

* Also co-occurred with the same to verbs in Bybee & Eddington's (2006) data from the 1900's

The situation with *poner(se)* + ADJ and *quedar(se)* + ADJ has some similarities to the one described between *fincar(se)* + ADJ and *quedar(se)* + ADJ in the previous

section. From her data, Balasch (2008) found 14 tokens distributed over 10 types in the 1600's and found 63 tokens distributed over 45 types in the 1800's. The increased number of tokens in her data is indicative of a rise in frequency of *poner(se)* + ADJ. A search for the verb *poner* occurring with an adjective on the *Corpus del español* (Davies 2002-) reveals that it went from occurring 0.25 times per 10k words in the 1600's to occurring 0.35 times per 10k words in the 1800's[13]. Taken one step further, there was most likely an increase in the overall frequency of *poner(se)* + ADJ at the same time there was a decrease in frequency of *quedar(se)* + ADJ. The fact that these two expressions of 'becoming' share so many adjectives, at the very least, could affect the probability that *poner(se)* + ADJ would be used instead of *quedar(se)* + ADJ with a set of shared adjectives. Considering that the other verbs of 'becoming' studied by Bybee & Eddington (2006) may have been emerging in usage sometime during the time span studied here, it is likely that their emergence may have also affected the drop in overall frequency of *quedar(se)* + ADJ from the 1600's to the 1900's.

4.2.1.4 Possible beginnings of *quedar(se)* + ADJ

Before investigating interaction with the expression of 'becoming' *fincar(se)* + ADJ, the adjective *solo* was deemed to be one of the key adjectives that played a role in how the verb *quedar(se)* came to be analyzed as a verb of 'becoming' based on analogy to usages that meant 'to remain'. Example (39) shows how *quedar(se)* is used with the adjective *solo* 'alone' to denote a change of state.

[13] These occurrences of *poner* used with an adjective on the *Corpus del español* (Davies 2002-) were not analyzed in detail to see if they all expressed a change of state. However, a brief scan indicates that the majority of the tokens of *poner* used with an adjective yielded by the search were tokens of the change-of-state expression *poner(se)* + ADJ.

(39) *E el conde quando vio que de otra manera no podia ser sino como queria el comun delos romeros no quiso ay quedar solo& fa zia lo mejor & cogio sus tiendas & fue se empos delos otros*.
'And when the count saw that there could be no other way than what the majority of the pilgrims to Rome wanted (it), (he) didn't want to be left alone and did his best and gathered his tents and went after the others.'
(*Gran conquista de Ultramar*, anon., 13th c.; Davies 2002)

This example typifies many of the 13th and some of the 15th century examples in the *fincar / quedar* study using Davies (2002) whereby the change of state is brought about by the movement of people away from the human subject of the construction. Viewed this way, if the count 'remains' he will find himself without his subjects, the pilgrims. By remaining he would undergo a change of state and be left alone and both meanings are present: 'remaining' and 'becoming'. Prior to this analysis, I proposed that it was possible that this ambiguity may have been one of the original paths that opened up the verb *quedar* to the analogical extension necessary for it to be used in a change-of-state construction.

Although this re-analysis of *quedar(se) solo* may have been influential in the increase in frequency of the verb *quedar* used to express a change of state, the fact that it is largely synonymous to the verb *fincar* must have also played a role. Because of the many synonymous meanings that the verbs shared, and the fact that *fincar* was much more frequent than *quedar* in the 1200's, there was already an established template by which a verb could mean both 'to remain' and 'to become'. As *fincar* experienced a sharp drop in frequency from the 1200's to the 1800's, without totally disappearing, it is plausible to imagine a process by which speakers attributed characteristics of *fincar* to *quedar* over time. One of these characteristics would be that it could function in a

change-of-state construction with many of the adjectives previously used in the construction *fincar* + ADJ.

4.3 Summary

This section has given an overview of the change-of-state construction *quedar(se)* + ADJ over time beginning with proposed categories (Section 4.1) and standardized frequency (Section 4.2.1). While most categories endured over the time span studied, the standardized overall frequency of *quedar(se)* + ADJ rose to a peak in the 1600's before dropping through the 1900's, forming a bell curve.

In order to include discussions of formulaicity, a prefab threshold and a threshold for conventionalized instances of constructions (CIC's) was proposed in Section 4.2.1.1. Part of the purpose of identifying CIC's is to show that there is a correlation between the overall standardized frequency and the portion of data above the CIC threshold. The main purpose in proposing a method for identifying prefabs is to show that these have longevity.

In order to examine reasons contributing to a bell curve of overall standardized frequency, other change-of-state verbs were examined. It appears that the rise in the overall frequency of *quedar(se)* + ADJ correlates to a decline in a similar verb of 'becoming', *fincar(se)* + ADJ as the former came to be used with adjectives previously associated with the latter (Section 4.2.1.2). Similarly, as proposed in Section 4.2.1.3, the decline in the standardized overall frequency of *quedar(se)* + ADJ may relate to the emergence of other verbs of 'becoming' as many adjectives came to be used with them.

Finally, this Section 4.2.1.4 explores some of the possible pathways by which a verb that denotes ‘remaining’, such as *quedar(se)* or *fincar(se)*, could be reanalyzed as a verb that denotes ‘becoming’.

Chapter 5: Exemplar clusters and analysis

This chapter presents a detailed analysis of four sets of proposed exemplar clusters of adjectives found in the change-of-state construction *quedar(se)* + ADJ in a written corpus from the 1200's to the 1800's: (a) the *solo* 'alone' clusters, (b) the *libre* 'free' clusters, (c) the *rico / pobre* 'rich / poor' clusters, and (d) the *alegre / satisfecho* 'happy / satisfied' clusters[14]. As mentioned in the previous chapter, the organization of adjectives into exemplar clusters was based on an analysis of each token, taking into account the context of the token and the exemplar clusters from surrounding centuries. The structure of the exemplar clusters proposed in Bybee & Eddington (2006) was a point of departure for the present clusters. Because the goal of this dissertation is to use the exemplar model to account for adjective types used in *quedar(se)* + ADJ over time, each set of clusters is presented diachronically in its own section to show how it has manifested in each century. The *solo* clusters and the *libre* clusters are the first two clusters analyzed, as they show how categories of adjectives endure and show coherency over many centuries. The *rico / pobre* clusters have less coherency but provide an opportunity for an analysis oriented toward concept of opposites in a category. Finally, the *alegre / satisfecho* clusters demonstrate how chains of family resemblance can link seemingly unrelated types.

5.1 Preliminary considerations: central members and the exemplar model

Central members of categories are also chosen based mainly on token frequency. In examining a group of similar adjectives, the one that had the highest token frequency for a given century is deemed to be the central member of those related types. However,

[14] Other clusters that are not discussed here are included in Appendix 1.

sometimes there are exceptions to the frequency criterion, especially in cases that appear idiosyncratic when looking at the same form in surrounding centuries. In this case, other factors may be considered. One of these factors is semantic generality. More general types tend to be central members; because of their generality they can be applied to a variety of contexts and, therefore, tend to appear with higher frequency. Furthermore, they are semantically related to a larger body of types than less general types. Because they are related to a larger body of types, general types may serve as flexible models of extension that can fan out in many directions. Another consideration in choosing the central member is endurance over several centuries. Finally, frequency data regarding *quedar(se)* + ADJ from the *Corpus del español* (Davies 2002-) is used as additional support. All of these factors will be discussed in the upcoming analysis when they are relevant.

One of the advantages of the exemplar model of representation is that it allows for categories with 'fuzzy' boundaries (Bybee & Eddington 2006). This eliminates the need to propose discrete patterns of language use in which categories do not interact, and in which category members may not appear simultaneously in other categories. In Bybee & Eddington's (2006) study, this was important to their analysis as they found that some adjective types appeared with more than one verb of 'becoming'. A flexible model, such as the exemplar model, could more accurately account for adjective distribution found in their data while including notions of the prototype model, such as chains of family resemblance. Sometimes these chains can link types that would otherwise be unrelated.

Because the adjectives in these data were organized into categories, it is tempting to view these as discrete and unrelated to one another. However, in a broader, more

holistic view, it is possible to see how many of the proposed categories are actually related. The figures presented in this section show how some central category members from the analysis of adjectives that denote social states (analyzed in Section 5.2) may relate to one another, and how the central category members depicting mental states (analyzed Section 5.3) may relate to one another. In other words, there may be a direct semantic link across clusters. In order to depict a manner in which the exemplar categories related to one another, the data from the semantic similarity study (Ch. 6) was considered[15]. Central members of clusters are in larger type and more peripheral members are in smaller type font.

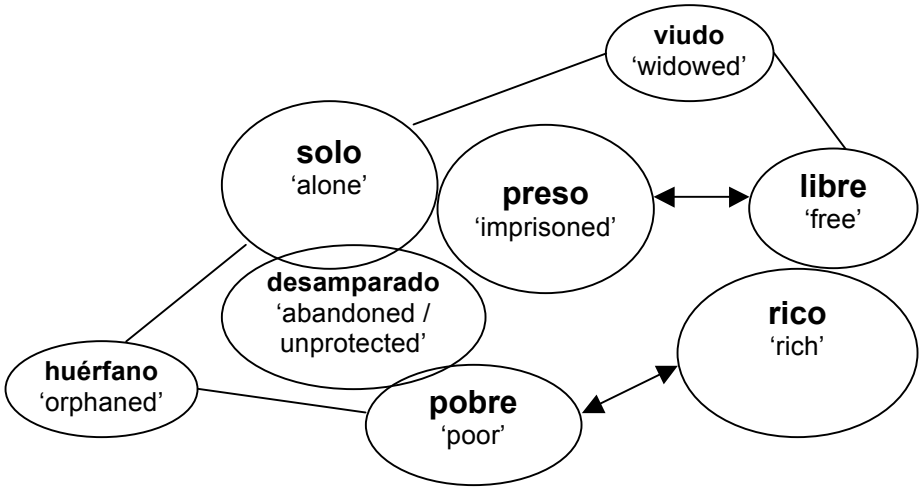


Figure 4. Chains of resemblance among categories portraying social states.

In the similarity study (Ch. 6), the type *solo* 'alone' was found to be related to types that portrayed 'aloneness' such as *viudo* 'widowed', *huérfano* 'orphaned', and *desamparado* 'abandoned / unprotected'. These types connect *solo* to others. The lines in Figure 4 connecting *solo*, *viudo*, and *libre* 'free' reflect the finding that *viudo* was rated

[15] These two figures (Figure 4 and Figure 5) were constructed after the process in which adjectives were analyzed and organized into proposed exemplar clusters. The MDS study was used to confirm the results of this analysis except in this case, where it is used to illustrate a theoretical point.

by speakers as highly similar to both of these types. There was another chain that linked *solo* to *libre*; *preso* 'imprisoned' was rated as similar to *solo* and to *desamparado*. Because opposites share some characteristics, *preso* is linked to *libre*, as indicated by the arrow. In a similar chain, *huérfano* was perceived as closely related to *solo* and *pobre* 'poor'. The type *desamparado* was rated as even more similar to *pobre* than *huérfano*. *Pobre* and *rico* are linked by virtue of being opposites. Finally, speakers perceived *rico* and *libre* as being highly similar.

Figure 5. Chains of resemblance among categories portraying mental / emotional states.

The types *suspenso* 'surprised' and *satisfecho* 'satisfied' are linked to one another through two different chains of types, as seen in Figure 5. First, *encantado* 'enchanted' appears in contexts that denote surprise and in ones that denote happiness. Second, the types *mudo* 'mute' and, to a greater degree, *inmóvil* 'motionless' were found to be perceived as relating to types that denoted calmness, such as *tranquilo* 'calm'. At the

same time, they were perceived as relating to types that denoted fear or surprise, such as *aterrado* 'terrified' and *suspenso* 'surprised'; a person can react to a fearful or surprising incident by becoming mute or motionless. *Inmóvil* also served as a link to *dormido* 'asleep' and *muerto* 'dead' and there was a strong link between *tranquilo* and *dormido.*

These two Figures (Figure 4 & Figure 5) show how seemingly unrelated types may be connected through a series of intervening links. This will be explored diachronically in section 5.3.1. Even though the sets of adjectives denoting social states and mental states were presented separately, there are links between them. For example, *triste* 'sad' could plausibly be perceived as similar to *solo* 'alone' and all of the types connected to it in Figure 4 based on the notion that it may be sad to be left alone, orphaned, or widowed (see Example (97), p. 180). Although there are many overlapping members linking exemplar clusters, the types in the *solo*, the *libre* and the *rico / pobre* clusters are presented separately for the ease of interpretation and presentation.

5.1.1 Issues regarding the generalizations *quedar(se) sin* + NOUN and *quedar(se) con* + NOUN

As mentioned in the previous section, *quedar(se) sin* + NOUN and *quedar(se) con* + NOUN are notational generalizations used in order to refer to a group of types. As there are many different types of prepositional phrases starting with *sin* 'without' and *con* 'with', it facilitates the discussion to use some generalizations when referring to these. These generalizations are not meant to be interpreted as actual exemplars, but as groups of exemplars united by their form. The different levels of generalizations are presented in Table 11, below.

Table 11. Overview of generalizations of *quedar(se) sin* + NOUN and *quedar(se) con* + NOUN

Generalizations	
quedar(se) sin + NOUN	*quedar(se) con* + NOUN
Sub-generalizations	
quedar(se) sin + HUMAN (a) *quedar(se) sin* + OBJECT (b) *quedar(se) sin* + NON-PHYSICAL OBJECT(c)	*quedar(se) con* + HUMAN (d) *quedar(se) con* + OBJECT (e) *quedar(se) con* + NON-PHYSICAL OBJECT (f) *quedar(se) con* + BODYNOUN + ADJ (g)
Exemplar Types	
quedar(se) sin padre (a) 'to be left without a father' *quedar(se) sin dinero*(b) 'to be left without money' *quedar(se) sin libertad* (c) 'to be left without freedom'	*quedar(se) con hijo* (d) 'to be left with a child' *quedar(se) con el reino* (e) 'to be left with the kingdom' *quedar(se) con honra* (f) 'to be left with honor' *quedar(se) con el cuello tuerto* (g) 'to be left with the neck twisted'

The generalizations *quedar(se) sin* + NOUN and *quedar(se) con* + NOUN are the broadest ones and include the pertinent sub-generalizations, and the groups of exemplars that form these. It further facilitates the discussion to break the broader generalizations into sub-generalizations because these show observable patterns in the analysis. For example, the exemplar clusters included in the sub-generalization *quedar(se) sin* + HUMAN (e.g. *quedar(se) sin padre / madre / caudillo* 'to be left without father / mother / boss) are related to *quedar(se) solo* 'to be left alone' in that they depict a situation in which the subject is left without human company. On the other hand, exemplar clusters included in the sub-generalization *quedar(se) sin* + OBJECT (e.g. *quedar(se) sin dinero / abrigo* 'to be left without money / jacket) illustrate a change more similar to *quedar(se) pobre* 'to become poor' in that the subject is left without material goods. The other three, *quedar(se) sin* + NON-PHYSICAL OBJECT, *quedar(se) con* + NON-PHYSICAL OBJECT, and *quedar(se) con* + BODYNOUN + ADJ do not tend to exhibit exemplars that belong to one

particular set of clusters. The exemplar types listed in Table 11 are examples from the data of the respective sub-generalizations as indicated by the subscript letters in parentheses.

Exemplars pertaining to these generalizations are prevalent in the data. *Quedar(se) sin* + NOUN appears in every century in this data and there is one token in Bybee & Eddington's (2006) data from the 1900's. *Quedar(se) con* + NOUN appears in this data from the 1200's to the 1600's. As will be seen in the analysis, there is both a great degree of consistency in the nouns that can be inserted into this slot, and a propensity for novel usage. Furthermore, the actual forms *sin* and *con* have been used to extend the clusters in which they appear. For example, exemplars of *quedar(se) con* + OBJECT can express a change of state in which the subject obtains possession of an object (e.g. *con el reino* 'with the kingdom') and relate to the central member *quedar(se) rico* 'to get rich' because they indicate an accumulation of wealth. Also, the same form can express changes in which the subject loses possession of something (e.g. *con pérdida sin enmienda* 'with loss without measure') and relate more to the central member *quedar(se) pobre* 'to become poor'.

5.2 Social states

This section presents clusters formed by adjectives used in *quedar(se)* + ADJ to describe a change of social state. It begins with the *quedar(se) solo* 'to be left alone' clusters in section 5.2.1. These have had both relatively high token frequency and endurance in the data, spanning from the 1200's to the 1800's. The central member, *solo*, was also one of the adjective types with the highest token frequency in Bybee & Eddington's (2006) study of data from the 1900's. This presentation is followed by the

libre ‘free’ clusters in Section 5.2.2, and the *pobre/rico* ‘poor/rich’ clusters in Section 5.2.3. Other adjectives and clusters that are used in *quedar(se)* + ADJ to illustrate a change of social state not included in the ensuing analysis are shown in Appendix 1, Section 1.

5.2.1 Being left alone: the *solo* ‘alone’ clusters

The *solo* ‘alone’ clusters figure prominently in this study because they document one of the key concepts in this diachronic study: prefabs have longevity. The prefab *quedar(se) solo* appears above the prefab threshold from the 1500’s to the 1800’s (see Table 12, p. 124) and above the CIC threshold from the 1200’s to the 1800’s[16]. One of the other things that this section demonstrates is that the proposed *solo* clusters are highly similar to each other from one century to the next during the entire span of time from which data was collected. *Quedar(se) solo* is always the central member and is consistently surrounded by similar, if not the same, types. The similarity study presented in Chapter 6 supports the proposed *solo* clusters by showing that these types are perceived as similar to one another. Also, the similarity study supports the notion that *quedar(se) solo* is the central member of these clusters.

On a semantic level, the common link among adjective types in these clusters is that there is a change of state in which the subject is left without the company of another person or group. Because this implies a change that the subject has in relation to other members of their social group, the *solo* clusters are placed in the section on social states.

[16] The CIC measurement (twice the token / type ratio) is applied to all types in my own data while the prefab measurement is applied only to the types *quedar(se) solo* and *quedar(se) libre* using Davies’ (2002-) data: verb + adjective sequences that account for at least 1% of all adjectives and 4% of all occurrences of *quedar(se)* + ADJ.

The opposites appearing in the upcoming clusters appear by virtue of showing a change of state in which people are united in some fashion.

5.2.1.1 The *solo* clusters in the 1200's

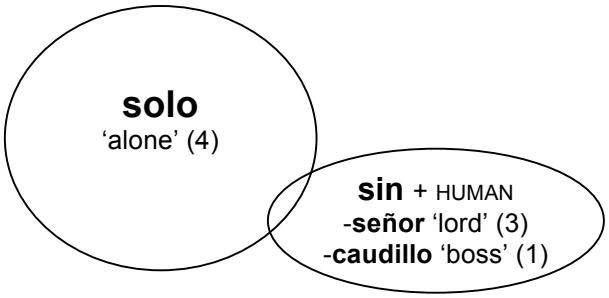


Figure 6, 1200's: The *solo* clusters.

The 1200's is the first century from which data was collected for this study. As indicated in Table 4 (Chapter 4:, Sect. 4.2.1), this century shows the lowest standardized overall frequency for the construction *quedar(se)* + ADJ: 0.42 occurrences per 10k words. Despite the low standardized frequency of the general construction, *quedar(se) solo* appears above the proposed CIC threshold, showing evidence of conventionalization. The central member of these clusters is argued to be *quedar(se) solo* from the 1200's through the 1800's, and even into the 1900's (Bybee & Eddington 2006).

Figure 6 shows the adjective types positioned based on their proposed semantic similarity to *solo*, illustrated by the intersection of the *solo* bubble and the *sin* + HUMAN bubble. The number of tokens found in the data is given in parentheses. This notational convention will be followed throughout the rest of this study.

One of the other important points regarding *solo* in this century is that all four tokens demonstrate an inherent ambiguity in which the expression *quedar(se) solo*

conveys both a sense of 'remaining' and 'becoming'. There is a clear change of state that is indicated, but this change only comes about by the movement of other animate entities. Consider Example (40) below:

(40) *E quando los de alema[ñ]a vieron esto metieronse al camino para tornarse a su tierra a pesar del emperador que aquellos que quedauan no querian morir como morian los otros: & quando el emperador aquello vio no quiso quedar solo & fuesse con ellos muy sa[ñ]udo & con gran pesar porque hauia muy bien comenado a passar.*
'And when those from Germany saw this they got on the road to return to their land, despite the Emperor, because those that stayed didn't want to die like the others died. And when the Emperor saw that, he didn't want to be left alone and he went with the others, irritated and with great despair because he had begun to enjoy himself very much.' (*Gran Conquista de Ultramar*. 13th c. Anon. O'Neill 1999)

The movement of the "others" could bring about a change of state in which the Emperor would be left alone, should he choose to remain. As discussed previously (Chapter 3:, Sect. 3.2), this ambiguity may have contributed to the reanalysis of a verb that expresses 'remaining', such as *quedar*, or its predecessor *fincar*, as a verb that additionally expresses 'becoming'. As time went on, *quedar(se) solo* became less ambiguous and was used unambiguously to express a change of state. In addition, Example (40) shows an occurrence of *quedar* meaning 'to remain / to stay' (e.g. *aquellos que quedauan* 'those who remained') pointing to the fact that this verb could be used to express 'remaining' even in proximity of a token of *quedar(se)* + ADJ that expresses a change of state.

In a similar vein, the types pertaining to *quedar(se) sin* + HUMAN express a change that happened to the subject through the actions of another human entity. However, these are not always as ambiguous as *quedar(se) solo* in this century. In Example (41), below, the people are left without a lord, but the change does not occur because the people wished to remain in a certain place or state of being. Instead, it is a

change of state that the people underwent because of the imprisonment of their lord by the authorities of Halapa.

(41) *& anduuieron buscando le a todas partes: & como no pudieron saber del tornaronse para su tierra & contaron aquella desauentura como hauian perdido su seor & no sabian como ni en que manera. E despues que lo supieron por la tierra fizieron muy grandes llantos porque quedauan como sin se[ñ]or. E a pocos dias supieron como era preso enla cibdad de halapa.*
'And they went looking for him all over. And since they couldn't find out anything about him they returned to their land and recounted that bad fortune, how they had lost their lord and didn't know how or in what way. And after it was known throughout the land there was great sobbing because they were left as though they were without a lord. And in a few days they found out how he was a prisoner in the city of Halapa. (*Gran Conquista de Ultramar*. 13th c. Anon. O'Neill 1999)

5.2.1.2 The *solo* clusters in the 1300's

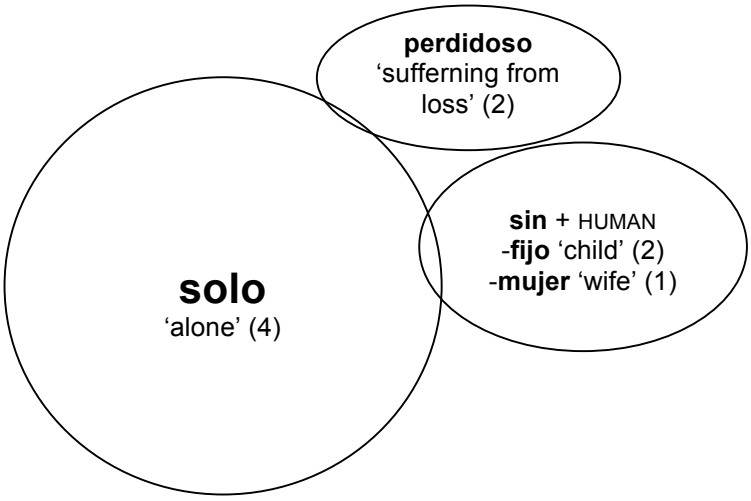


Figure 7, 1300's: The *solo* clusters.

As the standardized frequency of *quedar(se)* + ADJ more than doubles in the data during this century (0.96 tokens per 10k words) the central member *solo*, with 4 occurrences, appears at the proposed CIC threshold. A new related type, *perdidoso*

'suffering from (human) loss' appears in this cluster, as seen in Figure 7; Example (42) shows the context in which *perdidoso* appears.

(42) *[M]uy perdidosos quedaron los troyanos por la muerte dela rreyna*
The Trojans were left suffering from loss because of the death of the Queen. (*Sumas de la historia troyana*. 14th c. Leomarte. O'Neill 1999)

The DRAE (2006) gives the following definition for *perdidoso: adj. Que pierde o padece una pérdida* 'adj. That loses or suffers from a loss'. Used in *quedar(se)* + ADJ, this adjective specifies a change of state in which the subject undergoes a loss, in this case a loss caused by the death of the Queen, and implicates the negative emotional state that results.

The generalization *quedar(se) sin* + HUMAN shows its flexibility by extending to two human noun types that were not seen in the data in the previous century: *fijo* 'child' (*hijo* in Modern Spanish), and *muger* 'woman/wife' (*mujer* in Modern Spanish) as seen in Example (43) below.

(43) *E el rrey thereo non las pudo alcanc'ar que tan espantado quedo que en esa muy grand pyec'a non sopo a donde estaua / E quando acordo asy lo sopieron ellas agujsar que Jamas non las pudo aver E fueronse & pasaron la mar E quedo el rrey thereo syn fijo*$_1$*. E syn muger*$_2$ *muy triste*$_3$ *& envergonnado*$_4$ *E perdidoso*$_5$.
And King Thereo could not reach them so he became so scared that he didn't know where he was in that great room. And when he remembered that they knew how to take care of themselves he knew that he could never have them. And they left and crossed the sea. And King Thereo was left without a child$_1$. And without a woman$_2$, very sad$_3$ and embarrassed$_4$. And suffering from the loss$_5$. (*Sumas de la historia troyana*. 14th c. Leomarte. O'Neill 1999)[17]

This example shows six different tokens of *quedar(se)* + ADJ. The first is *tan espantado quedo* 'he became so scared' (analyzed in Appendix 1, Section 3). The other

[17] The types *quedar(se) triste* and *quedar(se) envergonnado* were placed in the *triste* clusters (Appendix 1, Section 2)

tokens come from a single occurrence of the verb *quedar(se)* in its 3s preterit form and the ensuing sequence of five of adjectives, indicated by the numbers in subscript. In this case, the open slot is filled with multiple adjectives all activated by the single occurrence of *quedar(se)*. These occurrences of *quedar(se)* + ADJ with multiple adjectives are common in the data and give insight into semantic commonalities perceived by the writer.

The change of state expressed by *E quedo el rrey thereo syn fijo* 'And King Thereo was left without a child', in Example (43) is an example of a change brought about by the movement of others. The king is left without his child and his wife through some sort of flight. It is a case of ambiguity in which the two senses 'remaining' and 'becoming' are present. However, it is presumed that the other adjectives that follow are less ambiguous because they result from the change caused by being left behind. Out of these adjectives, two, *syn fijo* and *perdidoso*, are proposed to belong to the *solo* clusters while *triste* and *avergonnado*, are proposed to belong to the *triste* 'sad' clusters shown in Appendix 1, Section 2.

5.2.1.3 The *solo* clusters in the 1400's

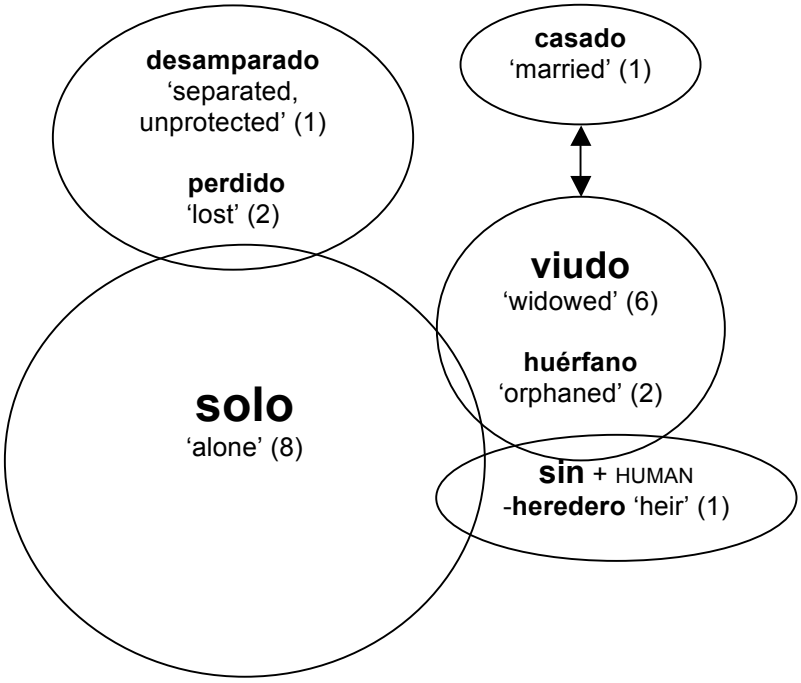


Figure 8, 1400's: The *solo* clusters.

In the data from the 1400's, there is another rise in standardized frequency as the construction *quedar(se)* + ADJ appears 1.18 times per 10k words. The central member *solo* occurs 8 times in the data, which is twice the proposed CIC threshold for this century. The generalization *sin* + HUMAN persists in this century with one type, *sin heredero* 'without heir'. There are some new types seen in the 1400's that reflect changes in the presence or absence of other humans with respect to the subject: *desamparado* 'separated, unprotected', *perdido* 'lost', *viudo* 'widowed', *huérfano* 'orphaned', and the opposite *casado* 'married'. In the case of *perdido*, *viudo*, and *huérfano*, the similarity between these types and *solo* is supported in the similarity study described in Chapter 6.

The type *quedar(se) viudo* 'to be left widowed' is included in the *solo* clusters because it is used in the construction *quedar(se)* + ADJ to describe a change, namely a

death, in which the subject finds themselves without a spouse as in Example (44). Also note that this example of *quedar(se)* + ADJ has two other adjectival predicates: *atribulada* 'sorrowful', and *triste* 'sad'.

(44) *Lo qual creyeron los thebanos haver acahescido por ira de los dioses, que vengaron la injuria de la divina majestad. E Níobes, quedando viuda y muy atribulada y triste de tantas muertes, vino en tan grande y durable silencio que parescía más una piedra que no se mueve que mujer.*
'Which the Thebanosites believed had happened because of the anger of the gods that avenged the insult to the divine majesty. And Niobe, becoming widowed, very sorrowful, and sad from so many deaths, entered into such a great and durable silence that she seemed more like a rock that doesn't move than a woman.' (*De las mujeres illustres en romance*. 15th c. (1494). Boccaccio, J. Lemir)

Becoming orphaned is a similar change to becoming widowed, as shown in example (45):

(45) *El vuestro maestro & fiel pastor se partio de nos & nos quedamos asi huerfanos mas yo espero firme mente que ayna lo veremos conusco.*
'Your teacher and faithful shepherd has parted from us and, because of that, we have become orphaned. But I firmly hope that soon we will see him with us.' (*Meditaciones.*15th c. *Pseudo-Augustine*. BVMC)

The two adjectives *perdido* 'lost', and *desamparado* 'separated, unprotected' are included in these clusters because, used in *quedar(se)* + ADJ, they both convey changes in respect to the presence of other members of the subject's social group. Example (46) is an example of how the form *perdido* 'lost' has extended the category into a more metaphorical sense. The young girl may not be physically lost in some sort of unfamiliar territory, but is lost metaphorically in respect to social values.

(46) *Vna muy bonita moça; avnque queda agora perdida la pecadora, porque tenía a Celestina por madre e a Sempronio por el principal de sus amigos.*
A very pretty young girl; although the sinner has now become lost, because she had Celestina for a mother and Sempronio as her main friend. (La Celestina. 15th c. de Rojas, F. BVMC.)

The idea of getting married is included as an opposite in this century because it reflects a change in the presence of a member of the subject's social group. Instead of losing a spouse, as in *quedar(se) viudo* 'to become widowed', the subject gains a spouse. In Example (47) the subject gets married, thereby undergoing a change in social state in which she gains a husband, and also becomes a countess[18].

(47) *E desque el Rey & la Reyna lo supieron pesoles de muerte & afinco mas el Rey la prision a doña leonor por eso & non la dexauan despues ver a ninguno de su valia % E enbiaronla presa a carmona & doña juana quedo condesa & casada con el conde.*
'And ever since the King and the Queen found out about it, it saddened them to death. But the King put Doña Leonor in prison for this and afterwards they didn't let her see any of her cohorts. And they sent her as prisoner to Carmona and Doña Juana became countess and married to the count.' (*Atalaya de las corónicas*. 15th c. Anon. O'Neil, 1999)

[18] The token *quedar(se) condesa* belongs to adjectives that denote social status shown in Appendix 1 (Section 1). It is included in the data because it is used here as an adjective. As Hopper and Thompson (1984: 714) observe, in discourse it is possible to manipulate word class, embedding nouns into adjectival formatives; this is accompanied by symptoms of 'decategorialization'. Evidence of this here is that there is no article and it is indistinguishable from the other adjective in this 'multiple', *casada* 'married'.

5.2.1.4 The *solo* clusters in the 1500's

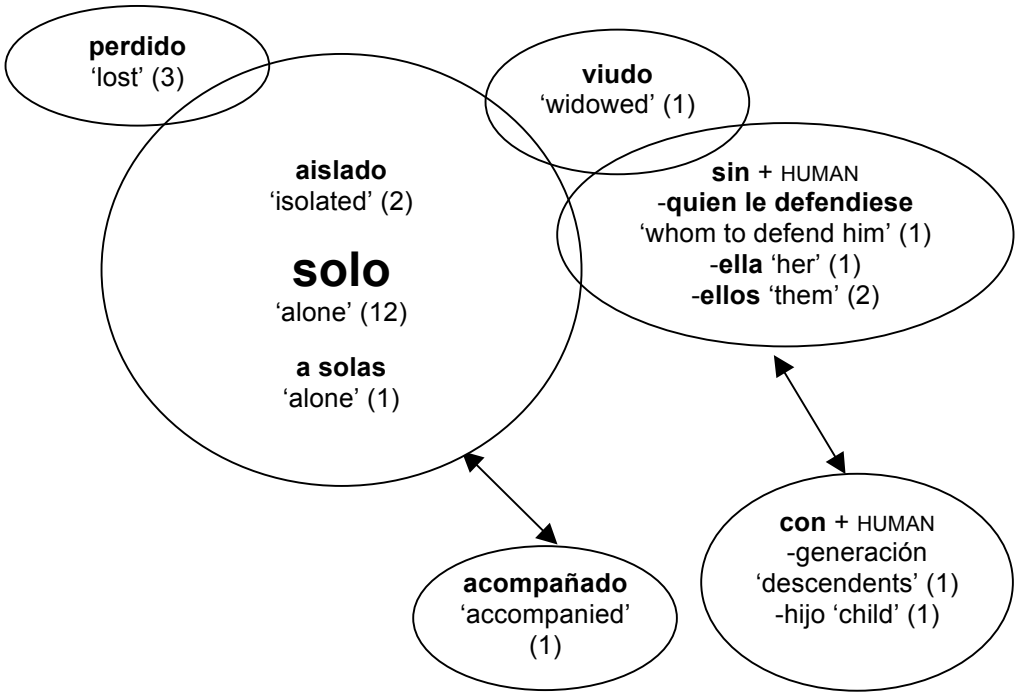


Figure 9, 1500's: Clusters centering on *solo.*

In the 1500's the standardized overall frequency of *quedar(se)* + ADJ more than quadruples, going from 1.18 to 4.85 occurrences per 10k words. With 12 occurrences in the data, the central member *solo* 'alone' is the only type that exceeds the CIC threshold in this century. Also, this is the first century in which *quedar(se) solo* appears above the proposed prefab threshold[19]. Despite the huge gain in overall frequency, both the adjective types and the overall shape of the proposed *quedar(se) solo* clusters are very similar to the ones in the previous century. Figure 9 shows that there are some new types including *aislado* 'isolated', *a solas* 'alone', and *acompañado* 'accompanied'.

[19] The sequence *quedar(se) solo* accouts for at least 1% of all occurrences of the adjective *solo*, and 4% of all occurrences of the construction *quedar(se)* + ADJ.

Quedar(se) solo continues to describe a change of state in which the subject is left alone based on the movement of other humans. Example (48) is an unambiguous change of state; Progne orders everyone to leave, causing a change of state in which she will be left alone with "your Highness".

(48) *TAURINO: Ya está todo aparejado, su Alteza bien puede entrar. PROGNE:No ha de ser d'essa manera. Solos hemos de quedar. ¡Salios todos afuera! TAURINO: ¡Pardiobre! Yo más quisiera ayudalles a maxcar. ¿Señora no quedaré a roer los ossezillos?*
'TAURINO: Everything is now suitable, your Highness may enter. PROGNE: It must not be this way. We must be left alone. Everyone go outside! TAURINO: By God! I just wished to help you all chew. Madame, shall I not stay to gnaw the bones?' (*Tragicomedia llamada Filomena*. 16th c. (1564). Timoneda, J. Lemir.)

There are still ambiguous tokens in this century in which the subject undergoes a change because they remain in one place while other group members move on to another destination as in the token of *quedar(se) sin* + HUMAN in Example (49). Here, because of his treacherous actions, the subject was left behind by the group with no one to defend him and a huge pile of money.

(49) *Prometióles mucho dinero si le bolvían sus vasos; bolviéronselos y no dio el dinero, por lo cual le dexaron, y quedó sin quien le defendiesse con grandes riquezas.*
'He promised them lots of money if they returned the glasses; They returned them to him and he didn't give the money, for which they left him, and he was left without anyone who would defend him and with great riches.' (Joan Timoneda, Valencia. 1569. *Sobremesa y alivio de caminantes*. Lemir. 2003)

A member of the generalization *quedar(se) con* + HUMAN appears in the *solo* clusters by virtue of expressing an opposite concept to *quedar(se) solo*. Example (50) shows how the subject, Noemí, underwent a change of state by getting married and coming out with a child and lineage in addition to estate and honor.

(50) *Avisada y discreta se mostró Noemí, suegra de Rut, en darle consejo que fuesse a espigar al campo de Booz Patriarca, y el modo que tuvo con él hasta ganarle por marido y quedar con honra y hazienda, y después con hijo y generación, que era lo muy preciado entre los hebreos.*
'Noemí, mother in law of Rut, showed herself to be informed and discreet in giving advice that he should go to collect the barley ears from the field of Booz Patriarch, and in the way she had with him until winning him as a husband and coming out with honor and estate, and afterward with a child and lineage, which was highly valued among the Hebrews.' (*Fructus Sanctorum y Quinta Parte del Flos Sanctorum*. 16th c. (1594). de Villegas, A. Lemir.)

5.2.1.5 The *solo* clusters in the 1600's

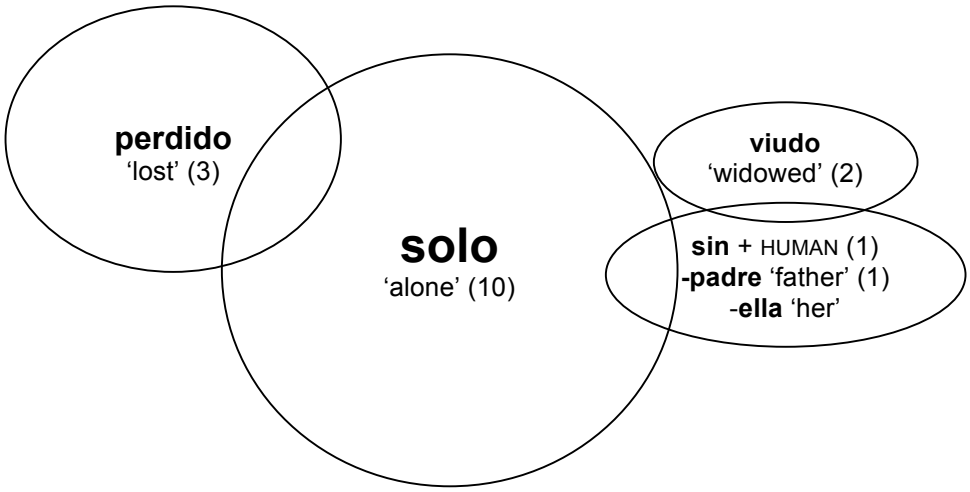


Figure 10, 1600's: The *solo* clusters.

The clusters centering on *solo* in this century show a great deal of continuity with the ones from the previous century. With the construction *quedar(se)* + ADJ occurring at a standardized overall frequency of 6 times per 10k words in the data, this is the century with the highest standardized frequency of them all for the general construction. As in the 1500's, the central member *solo* is the only one from these clusters that occurs above the CIC threshold of 4 for this century and it continues to appear above the prefab threshold.

The data show that in this century, *quedar(se) solo* 'to be left alone' does not always depend entirely on the movements of others in relation to the subject. In Example (51), the two subjects find themselves left alone through their own actions of fighting. This example comes from a play and is actual stage directions for the actors. It can be assumed that these actors would be fighting all around the stage until they came to a space that was free of other actors. Instead of being left alone by remaining in the same place as other humans move away, these actors would presumably move away to a space where they find themselves alone.

(51) '*Van peleando, entrando y saliendo, hasta que quedando solos don Gonzalo y Menalipe dicen:...'*
'They go fighting, coming in and going out, until being left alone, Don Gonzalo and Menalipe say:…' (*Amazonas en las Indias*. 17th c. (1635). de Vega L. BVMC.)

In Example (52), the adjective *perdida* is used to describe a state that the subject would enter into if she were to leave her husband. As in Example (46), from Section 5.2.1.3 above, it describes a state of metaphorical separation from a social entity instead of a state of being physically lost.

(52) *Todo era afligirme, todo llorar y todo dar a don Manuel quejas; unas veces, con caricias, y otras con despegos, determinándome tal vez a dejarle y no tratar más de esto, aunque me quedase perdida.*
'Everything was me grieving, crying and giving complaints to Don Manuel; sometimes, with caresses, and others with disinterest, making up my mind perhaps to leave him and not deal with this anymore, even though I may become lost.' (*El castigo de la miseria*. 17th c. Zayas y Sotomayor, M. BVMC.)

5.2.1.6 The *solo* clusters in the 1700's

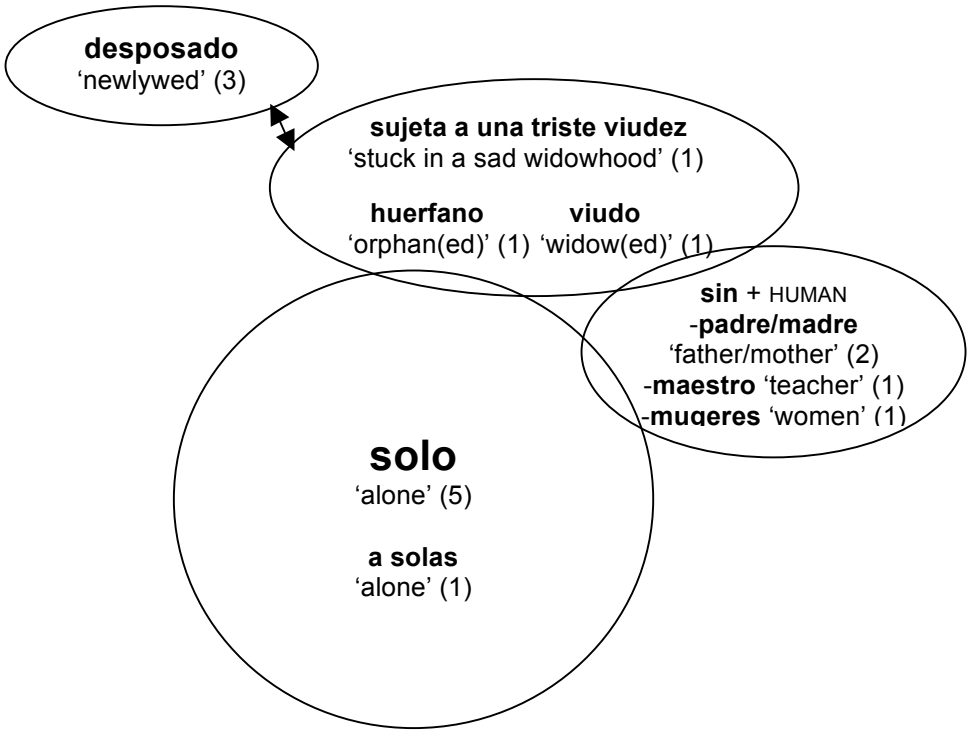


Figure 11, 1700's: The *solo* clusters.

The construction *quedar(se)* + ADJ maintains a relatively high standardized frequency in this century at 5.82 occurrences per 10k words in these data. Additionally, *quedar(se) solo* continues to appear above the CIC threshold.

The novel usage *sujeta a una triste viudez* 'stuck in a sad widowhood', seen in Example (53) below, is included in these clusters because it portrays a change of state in which the subject enters widowhood.

(53) *y después que, asegurados todos de la verdad del caso por muchas verídicas cartas y no menos verdaderos testigos, quedó mi desconsolada madre sujeta a una triste viudez.*
'and after which, all assured by the truth of the matter by many truthful letters and no less truthful witnesses, my disconsolate mother became stuck in a sad widowhood.' (*Los trabajos de Narciso y Filomela*. 18th c. Martínez Colomer, V. BVMC.)

The opposite of a change that leaves the subject without a family member would be one in which the subject gains a family member, as demonstrated by the adjective type *desposados* 'married' in Example (54).

(54) *Narciso y Filomela, después de haber dado rendidas gracias al cielo que les libró de tantos riesgos y peligros, quedaron solemnemente desposados con aplauso general de ambos reinos.*
'Narciso and Filomela, after having given devoted thanks to heaven that freed them from so many risks and dangers, became solemnly married with the general applause of both kingdoms.' (*Los trabajos de Narciso y Filomela*. 18th c. Martínez Colomer, V. BVMC.)

The adjective type *a solas* 'alone' appears again in the data. It is synonymous in its usage to *solo*. As shown in Example (55), this token represents a change of state in which the subject finds themselves alone through the movement of other humans.

(55) *Viendo mis padres y mis amigos que sus consolatorias persuasiones no eran de provecho, hicieron lo que les supliqué, que era dejarme solo en mi propia estancia, por ver si partiría algunas treguas con mis congojas. Quedé a solas sin más compañía que la de mis desgracias.*
'My parents and friends, upon seeing that their consoling persuasions were not benefiting, did what I begged them, which was to leave me alone in my own place, to see if I would break away from my anguish. I was left by myself without any company other than that of my misfortunes.' (*Los trabajos de Narciso y Filomela*. 18th c. Martínez Colomer, V. BVMC.)

5.2.1.7 The *solo* clusters in the 1800's

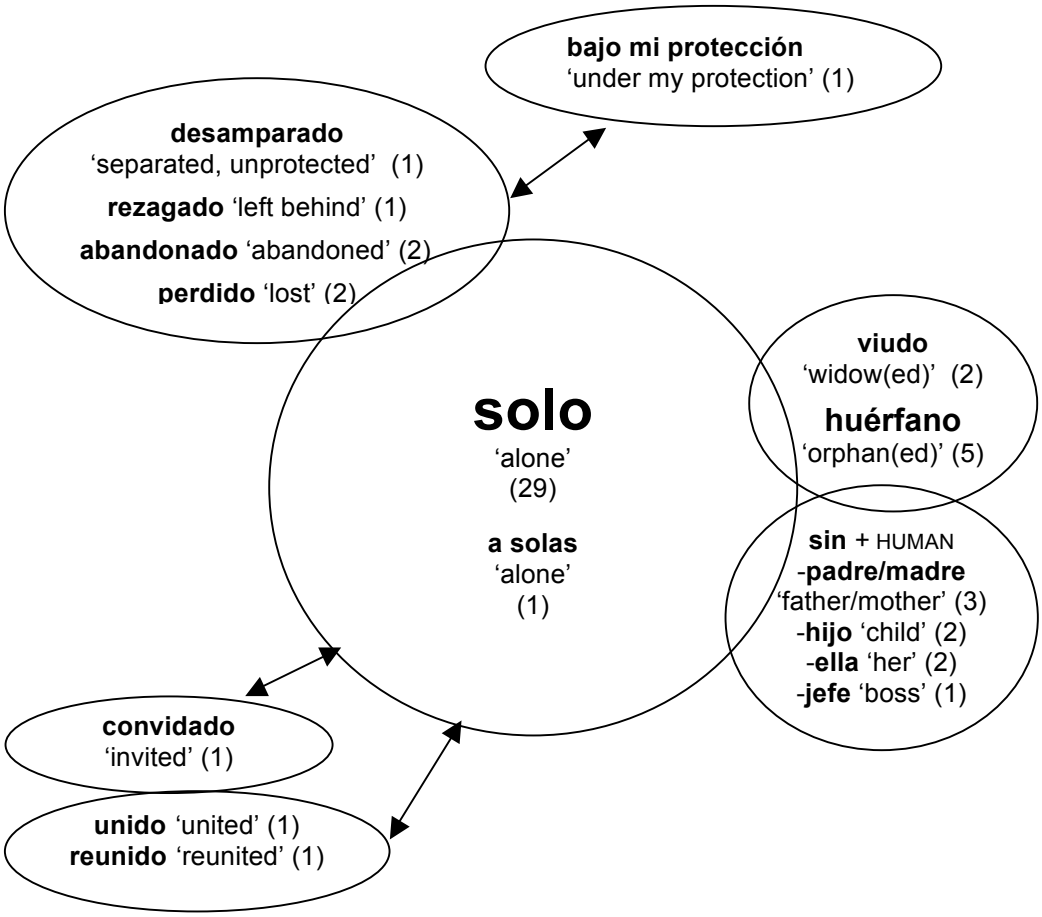


Figure 12, 1800's: The *solo* clusters.

In the 1800's, there is a increase in type frequency in the *solo* clusters and an increase in the relative and standardized frequency of *quedar(se) solo* 'to be left alone'; it accounts for 11.5% of all tokens and has a standardized frequency of 4.13 occurrences per 10k words in these data. As would be expected, it continues to appear above the CIC threshold. There are more detailed comments on this in the ensuing overview of *quedar(se) solo* presented in Section 5.2.1.8.

Many of the same types have persisted such as: *perdido* 'lost', *viudo* 'widowed', and *huérfano* 'orphaned'. All of the types pertaining to *sin* + HUMAN have appeared previously in the data except for *sin jefe* 'without boss', and even this has related precedents (e.g. *sin caudillo* 'without a boss' in the 1200's). *Desamparado* 'separated, unprotected' appears again after not being seen in the data since the 1400's. There are also new opposites: *unido* 'united', *reunido* 'reunited', *convidado* 'invited' and *bajo mi proteccion* 'under my protection'.

The adjective *desamparado* 'separated, unprotected' appears again in the data in a 'multiple' with the adjective *huérfano* 'orphaned' and emphasizes a sense of separation instead of a lack of protection. In example (56), the two tandem adjectives emphatically convey that the subject will be separated from highly regarded ideals.

(56) *Quedarás huérfano y desamparado de ideales que te sublimen y ennoblezcan, algo más que las absurdas abstracciones metafísicas con que hoy te engañan.*
'You shall become orphaned and separated from ideals that aggrandize and ennoble you, somewhat more than the absurd metaphysical abstractions with which they fool you these days.' (José María de Pereda. 1888. *Esbozos y rasguños*. 19th ctry. Biblioteca Virtual Miguel de Cervantes, 1999.)

For the first time in the data, the adjective *abandonado* 'abandoned' appears. Example (57) shows how the subject is abandoned by an implied person, or group, to be left at the mercy of another. Even though it is not used in its adjectival form, the noun *desamparo* 'abandonment' appears in the situation described and contributes to the sense of desolation that is portrayed.

(57) *Voy nuevamente a quedar abandonada a merced de un hermano caprichoso, y a todo el horror de un desamparo eterno.*

'I shall become newly abandoned at the mercy of a capricious brother, and to all the horror of an eternal abandonment.' (*Alfredo; Drama trágico en cinco actos*. 19th c. Pacheco, J.G. BVMC.)

The idea of being invited (*convidado*) to a great social event such as a wedding and baptism, as seen in Example (58), is proposed to be an opposite to *solo.* This is based on the idea that the subject should join a social entity instead of remaining separate.

(58) *Conque, con Dios, don Plácido; queda usted convidado para la boda y para el bautizo.*
'So, with God, Don Plácido; you are invited to the wedding and to the baptism.' (*A fuerza de arrastrarse*. 19th c. Echegaray, J. BVMC.)

Similarly, the idea of being united (*unidas*) implies that there is a joining of social entities instead of an ongoing separation. In the short but poignant Example (59), the souls of the subjects have been united forever instead of being left alone forever.

(59) *Me vio, le vi, y nuestras almas quedaron unidas para siempre.*
'He saw me, I saw him, and our souls became united forever.' (*Ángela*. 19th ctry. Tamayo y Baus, M. BVMC.)

5.2.1.8 Overview of clusters centering on *solo*

One of the most important findings for *quedar(se) solo* is the fact that it has appeared in these data above the CIC threshold in each of the seven centuries studied (1200's – 1800's), and appeared above the prefab threshold from the 1500's to the 1800's (see Table 12, below). This provides strong evidence that prefabs have longevity; *quedar(se) solo* has been a prefab for at least 500 years and has shown evidence of conventionalization for 800. Furthermore, analysis of the *quedar(se) solo* clusters shows that the proposed category has been consistent throughout the data studied. The adjective type *solo* has been present in the data from the onset and has been the proposed central member of these clusters throughout the span of the investigation. It has not shifted from

one adjective type to another, as has been the case in other sets of exemplar clusters (e.g. the *rico / pobre* clusters in Section 5.2.3 and the *alegre / satisfecho* clusters in Section 5.3.1). This set of clusters is coherent through time and the surrounding types have been consistent, as seen in Figures 6 through 12 (also, see Table 15, p. 129). This has lead to a similarity of proposed category structures from century to century.

In order to illustrate the method used in arriving at a proposed prefab threshold (described in Chapter 4:, Section 4.2.1.1), Table 12 shows the occurrences per million[20] words of the adjective *solo* 'alone' in the first column, the percentage of all of these occurrences of *solo* accounted for by *quedar(se) solo* in the second column, the occurrences per million words of *quedar(se)* + ADJ in the third column, and the percentage of *quedar(se)* + ADJ accounted for by *quedar(se) solo* in the fourth column which also provides the occurrences per million. The final column indicates the centuries in which *quedar(se) solo* appeared above the prefab threshold. All figures come from the *Corpus del español* (Davies 2002-) and there was no intervening linguistic material between the verb and adjective in any of the tokens. From the 1200's to the 1400's *quedar(se) solo* does not meet both of the frequency criteria of accounting for at least 4% of all occurrences of *quedar(se)* + ADJ, and at least 1% of all occurrences of the adjective *solo* in Davies' (2002-) data. However, it occurs at above the prefab threshold from the 1500's to the 1800's. We see that by the 1800's, even though there are signs of a drop in the overall frequency of the construction *quedar(se)* + ADJ, *quedar(se) solo* expands, accounting for 3% of all occurrences of the adjective *solo* and for 11.6% of all occurrences of *quedar(se)* + ADJ.

[20] Occurrences per million are used since the data came from Davies' (2002-) corpus which uses this denominator for frequency measurements.

Table 12. Figures used in calculating prefab threshold of *quedar(se) solo* (Davies 2002-)

	# *solo* per M	# *quedar(se) solo* (%) / *solo*	# *quedar(se)* + ADJ per M	# *quedar(se) solo* (%) / *quedar(se)* + ADJ	Prefab status
1200's	156.8	0.26%	3.0	0.4 (13.3%)	
1300's	155.8	0.45%	25.5	0.7 (2.7%)	
1400's	351.9	0.34%	35.3	1.2 (3.4%)	
1500's	365.1	1.64%	109.2	6 (5.5%)	x
1600's	392.2	2.5%	115.8	9.9 (8.5%)	x
1700's	442.3	1.2%	111.2	5.3 (4.8%)	x
1800's	413.1	3%	108.4	12.6 (11.6%)	x

x *quedar(se) solo* surpassed the prefab threshold

Another way to view the token frequency of this type in relation to other types is to calculate the percentage of all *quedar(se) solo* tokens out of all other tokens. Table 13 shows the total number of tokens of *quedar(se)* + ADJ found in the database per century, gives the number of tokens of *quedar(se) solo*, the percentage of all tokens per century represented by *quedar(se) solo* in parentheses, the standardized overall frequency of *quedar(se)* + ADJ as it occurs per 10k words, the standardized overall frequency of *quedar(se) solo* per 10k words, and the standardized overall frequency of the adjective *solo* per 10k words.

Table 13. Number and percentage of tokens of *quedar(se) solo* per century

	Total # of tokens	# of tokens *quedar(se) solo* (% of total # of tokens)	# *quedar(se)* + ADJ per 10k words	# *quedar(se) solo* per 10k words	# *solo* per 10k words
1200's	43	4 (9.5%)	0.42	0.04	1.57
1300's	76	4 (5.5%)	0.96	0.05	1.56
1400's	196	13 (6.5%)	1.18	0.08	3.52
1500's	271	13 (5%)	4.85	0.23	3.65
1600's	280	10 (3.5%)	6.00	0.22	3.92
1700's	253	5 (2%)	5.82	0.11	4.42
1800's	255	29 (11.5%)	2.64	0.30	4.13
1900's*	181	21 (11.5%)	1.83	0.21	4.17

* Figures calculated from Bybee & Eddington (2006)

Looking at the percentages it appears that there is a correlation between the overall frequency of *quedar(se)* + ADJ and the token frequency of *quedar(se) solo* in the data; as the expression of 'becoming' becomes more frequent, *quedar(se) solo* represents a smaller percentage of the total tokens. The 1600's and the 1700's were the centuries with the highest overall frequency of *quedar(se)* + ADJ with 6 and 5.82 tokens per 10k words respectively and were the centuries in which tokens of *quedar(se) solo* represents the lowest percentage of all tokens. It is likely that the proportionally lower rate of occurrence of *solo* in the 1600's and the 1700's is partially due to the fact that there are many other CIC's that appear in the data, a situation that gives *solo* competition.

Another factor that must be figured into the lower standardized frequency of *quedar(se) solo* in the data from the 1600's and 1700's is the fact that the 1700's presents a peculiar case. Looking at the standardized frequency of *quedar(se) solo* as measured per 10k words, in the 1700's there is a dip in standardized frequency that seems conspicuous when considering surrounding centuries. In the 1500's *quedar(se) solo* appeared 0.23 times per 10k words, in the 1600's 0.22 per 10k words. However, in the 1700's it drops to 0.11 occurrences per 10k words before going up to 0.30 occurrences per 10k words in the 1800's. While this could be partially attributed to the dilution mentioned above, it is also likely that, by happenstance, the works chosen for this century had an uncommonly low number of occurrences *quedar(se) solo*. Another alternative is that it was simply less frequent for unknown reasons. A search on the Davies (2002-) corpus reveals a similar dip in frequency; *quedar(se) solo* occurred 6 times per million words in the 1500's, went up to 9.9 per million in the 1600's, dipped to 5.3 occurrences per million in the 1700's, and went up again this time to 12.6 occurrences per million in

the 1800's. This pattern is similar to the shifts in standardized overall frequency shown in my data.

This idiosyncratic dip in the 1700's does not manifest itself in the standardized frequency of the adjective *solo*, as seen in the final column of Table 13. Instead, the adjective *solo* reaches its highest overall standardized frequency in this century. In fact, the two centuries with the highest overall standardized frequency for *solo* are also the ones with the highest standardized frequency for *quedar(se)* + ADJ, the 1600's and the 1700's. However, for *quedar(se)* + ADJ the 1600's is the peak and for the adjective *solo* the 1700's is the peak.

Table 14. Number and percentage of types accounted for by *quedar(se) solo* per century

	# Types in *solo* clusters	Token / Type ratio	Total # types
1200's	3 (10.5%)	1.48	29
1300's	4 (10%)	1.9	40
1400's	7 (6%)	1.62	121
1500's	9 (6%)	1.77	153
1600's	4 (2.5%)	1.63	171
1700's	9 (6.5%)	1.87	135
1800's	16 (11.5%)	1.80	141
1900's*	6 (11%)	3.35	54

Table 14 shows the number of types in the *solo* clusters per century with the percentage of all types that these represent given in parentheses, the total number of types in the data per century, and the token / type ratio for each century. The types included here are the ones that directly intersected with *solo* which excludes types from the generalizations *quedar(se) sin* + NOUN and *con* + NOUN that are related only structurally to *quedar(se) sin* + HUMAN (as mentioned above in Section 5.2.1, p. 105). In the 1200's, 1300's, 1800's, and the 1900's (data from Bybee & Eddington 2006) the types found in the *solo* clusters account for ten percent or more of all types found in the data. The

aforementioned centuries bracket the time span from the 1400's to the 1700's in which the types pertaining to the *solo* clusters accounted for fewer than 10% of all types. This change in the percentage of types accounted for by the *solo* clusters may be attributable to some factors that will be explained here. Regarding the early years, the data available for the 1200's to the 1400's was scarcer than in subsequent centuries. Consequently, the number of tokens attained was fewer than 200 tokens in each of these centuries (see Table 2, p. 61, and Table 4, p. 75). This may not have as much of an impact in the 1400's since I was able to collect almost 200 tokens. However, in the 1200's and the 1300's, I was only able to find 43 and 75 tokens, respectively, in the available texts. Out of these, there were 29 types in the 1200's and 40 in the 1300's. The fact that types in the *solo* clusters account for 10% and 10.5% respectively of all types found in the data could indicate that the *solo* clusters represented a productive category that was established even in these early centuries; *quedar(se) solo* appears in every century in the data as do *quedar(se) sin* + HUMAN types. In the 1300's, the actual types found in *quedar(se) sin* + HUMAN are different from the previous century, but the idea of being left without a member of a social group persists. Also, in the 1300's the type *quedar(se) perdidoso* 'to be left suffering from loss' appears. This adjective is related in form and meaning to the type *perdido* 'lost' which is found in the data from the 1400's to the 1800's, excepting the 1700's in which it did not appear in the data.

In regards to the later centuries, specifically the 1800's, there is rise in the percentage of types found in the *solo* clusters as they account for 11.5% of all types found in the data. There is strong evidence that this relates to the notable rise in the standardized frequency of the central member, *quedar(se) solo*, in this century. It goes

from occurring 0.22 per 10k words in the 1600's, drops idiosyncratically to 0.11 per 10k words in the 1700's, and shoots up to occurring 0.30 times per 10k words in the 1800's (as shown in Table 13, above). The fact that there is a simultaneous gain in the standardized frequency of the central member *quedar(se) solo,* and the type frequency of the category to which it pertains provides evidence that there is a correlation between the two phenomena. As *quedar(se) solo* becomes more frequent in usage, so do the opportunities for speakers to discover and use analogous usages to this central member. This correlation between a gain in the token frequency of a member and an extension of related types will be seen in upcoming categories.

In order to support the hypothesis that categories can maintain cohesion over time Table 15 gives a timeline that shows types appearing in the *solo* clusters that appeared in the data in at least three centuries. In a span of 800 years, *quedar(se) solo* appeared in data as a CIC in all of them and above the proposed prefab threshold[21] in four (1500's – 1800's). The type *quedar(se) viudo* 'to be left widowed' appeared from the 1400's to the 1800's in my data, and in the 1900's in Davies (2002-). The generalization *quedar(se)sin* + HUMAN is given here to show that, even though the specific types change, the structural elements of this generalization have persisted. Beyond being merely a generalization, these types are related; they depict a situation in which the subject underwent a change in which he or she was left without another human, be it a king, boss, child, or parent. In the case of *quedar(se) huérfano* 'to be left orphaned', it is likely that it did appear in written works in the 1500's and 1600's. It is probably just a matter of happenstance that it did not

[21] The difference between CIC's and types that appear above the prefab threshold is a methodological difference. The CIC measure is easily applied to all types in these data while the proposed prefab threshold was only applied to *quedar(se) solo* and *quedar(se) libre* and was derived from Davies' (2002-) corpus. They are not mutually exclusive classifications. Refer back to 4.2.1.1 for more detail.

appear in these data; in the *Corpus del español* (Davies 2002-) there were occurrences of *quedar(se) huérfano* in both of these centuries. The same situation applies to *quedar(se) perdido* 'to become lost' and *quedar(se) sin él/ella* 'to be left without him/her' in the 1700's; there were occurrences of both of these in the *Corpus del español* (Davies 2002-) in the 1700's.

Table 15. Types from the *solo* clusters that appear in at least 3 centuries.

	1200	1300	1400	1500	1600	1700	1800	1900
solo 'alone'	#####	####	####	####	####	####	####	BBB
viudo 'widowed'			####	====	====	====	====	DDD
huérfano 'orphaned'			====	DDD	DDD	====	####	DDD
perdido 'lost'			====	====	####	DDD	====	
sin padre 'w/o father'				DDD	====	====	====	DDD
sin + HUMAN	====	====	====	====	====	====	====	BBB

(====) = presence of type in data below CIC threshold
(####) = presence of type in data above CIC threshold
(DDD) = presence only in Davies *Corpus del español* (2002-)
(BBB) = presence of tokens in Bybee & Eddington (2006)

Considering the time span represented by the data, it is important to note that formulaic language shows endurance in usage. This section has shown how a high frequency formulaic sequence *quedar(se) solo* has been the most plausible central member of an enduring category. Not only does the category itself persist diachronically, but there is evidence of a correlation between the expansion of the category (i.e. rise in type frequency) and the rise in standardized frequency of the central member, *quedar(se) solo*. The fact that this category expands even as the construction *quedar(se)* + ADJ shows evidence of waning (i.e. drop in standardized frequency from the 1600's onward) is evidence that the frequency of the central member is related to the number of types that appear in usage.

5.2.2 Brushes with authority: the *libre* 'free' clusters

The *libre* 'free' clusters display less coherency over time than the *solo* clusters but more than the *pobre / rico* clusters (Section 5.2.3, below). The type for which these clusters is named, *libre* 'free', does not appear in my data until the 1300's, which, at three occurrences, is just below the CIC threshold of four for this century. However, in the span from the 1400's to the 1800's, *libre* appears noticeably above the proposed prefab threshold, excepting the 1700's in which it idiosyncratically does not appear in my data. However, evidence from the *Corpus del español* (Davies 2002-) indicates that *libre* should still be considered in the analysis, which leads to a practice of coordination with this outside database described in Section 5.2.2.3. Like the *solo* cluster (but unlike the *rico / pobre* clusters, in which the central member was sometimes unclear, or changed from century to century, as we will see below), it is clear that there is one central member in the span from the 1300's to the 1800's, and that it is *libre*. Furthermore, using the methods described in above (Chapter 4:, Section 4.2.1.1), *quedar(se) libre* appeared above the prefab threshold from the 1300's to the 1800's, a span of seven centuries; this shows that prefabs have longevity. Because *libre* is the consistent central member, this set of clusters is referred to as the *libre* clusters.

As the title of this section indicates, *libre* 'free' and the other types proposed to appear in these clusters have to do with changes of state that are produced by having an authority figure exert their will over the subject. Consequently, these tend to reflect changes in the social status of the subject in respect to some type of governing body. A king may proclaim that a prisoner go free, be thrown in prison, or be obligated to carry out some task. Every set of clusters in this section has types that, used in *quedar(se)* + ADJ, are related to acquiring freedom (e.g. *libre* 'free', *ahorrado* 'spared', *perdonado*

'pardoned') and opposite types related to some kind of constraint on freedom (e.g. *preso* 'imprisoned', *obligado* 'obligated', *condenado* 'condemned').

5.2.2.1 The *libre* clusters in the 1200's

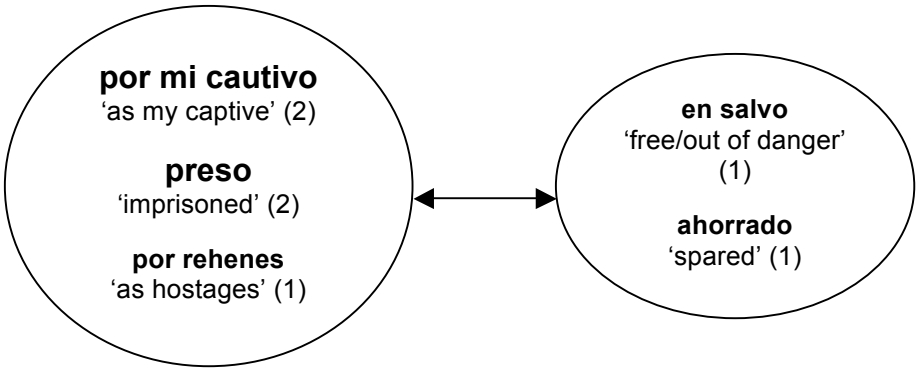


Figure 13, 1200's: The *libre* clusters.

In the 1200's, *libre* does not appear in the data despite the fact that, used in *quedar(se)* + ADJ, adjectives related to being set free do appear (e.g. *en salvo* 'free/out of danger', and *ahorrado* 'spared')[22]. In the opposite cluster, there is no clear central member as both *preso* 'imprisoned' and *por mi cautivo* 'as my captive' have two occurrences each in the data. No types occur with a token frequency above the CIC threshold. Nevertheless, *preso* is proposed to be the central member in this century since it is more general than *por mi cautivo*. Being 'imprisoned' is a change of state that could happen in a variety of settings whereas being left 'as my captive' refers to a becoming the prisoner of a specific individual.

In the *preso* 'imprisoned' cluster, all three adjective types used in *quedar(se)* + ADJ depict a change of state in which the subject became imprisoned, thereby losing freedom. In Example (60) the army host encloses the moors in what is presumably a

[22] There is one token of *libre* used with *quedar* in Davies' (2002-) data in the 1200's, but it does not indicate a change of state.

walled city typical of medieval times. Not only do the Moors become imprisoned, but many are killed by rocks and spears hurled down upon them by the host.

(60) *mas los dela hueste los encerraron dentro enla villa: assi que de encima delos muros & delas torres les dauan con piedras & con lanas de manera que quedaron muchos moros muertos & presos.*
'But those from the host enclosed them inside of the town: this way, from on top of the walls and from the towers, they hit them with rocks and with spears because of which many Moors were left dead and imprisoned.' (*Gran Conquista de Ultramar*. 13th c. Anon. O'Neill 1999)

In a manner similar to many tokens of *quedar(se) solo*, especially in the earlier centuries, there is a degree of ambiguity in which 'remaining' and 'becoming' are both present in the context. As has been detailed in Section 5.2.1 on *quedar(se) solo*, a subject may find themselves alone by remaining in one place as their fellow humans take leave. In the context of imprisonment, some people are allowed freedom by a governing body while those who remain become prisoners or hostages. In Example (61), after a peace accord, the sultan facilitates the free movement of the Christians by building them bridges and roads. Yet the king and the bishop, upon remaining with the sultan, undergo a change of state. Whatever their status was before, they are now hostages used to guarantee peace until Damiata has been turned over to the sultan.

(61) *E desque la paz fue otorgada dela vna parte & dela otra el soldan mando cerrar las acequias & fazer puentes & caladas por do los xpistianos pudiessen salir del agua a tierra seca: & despues dixo el soldan al rey que queria rehenes porque fuesse seguro dela paz fasta que le ouiessen entregado a damiata. E estonces el rey & el obispo quedaron por rehenes.*
'And ever since peace was granted from one part and the other, the sultan ordered that the ditches be closed, and to build bridges and roads from where the Christians could come out of the water onto dry land: And afterwards the sultan told the king that he wanted hostages so that he could be sure of peace until they had turned over Damiata to him. And then the king and the bishop remained as hostages.' (*Gran Conquista de Ultramar*. 13th c. Anon. O'Neill 1999)

The prepositional phrase *en salvo* 'free / out of danger' has connotations not only of being safe, but of being free from the clutches of authority. The DRAE gives the following definition of *en salvo*: *en libertad, en seguridad, libre de peligro* 'in freedom, in safety, free of danger'. In Example (62), King John finds out that the emperor wants to kill him and wishes to fool him in order to be left free of danger. As in the previous examples, the state of either being in danger, or becoming free of it, has to do with the subject's status in relation to a social entity with authority.

(62) *& quando el emperador oyo aquella razon fue mouido a tanto que los queria fazer matar: el rey juan que era en barlet supo el fecho & fue en muy gran cuydado porque era en medio dela tierra del emperador & penso en que manera podria enga[ñ]ar al emperador & quedar en saluo: & embio le vn mensajero que dixiesse que queria fablar conel.*
'And when the emperor heard the news he was so moved that he wanted to have them killed: King John, who was in Barlet, found out about this fact and went forth with great care because he was in the middle of the emperor's land, and he thought of a way that he could fool the emperor and be left free of danger: And he sent him a messenger to say that he wanted to speak with him.' (*Gran Conquista de Ultramar*. 13th c. Anon. O'Neill 1999)

5.2.2.2 The *libre* clusters in the 1300's

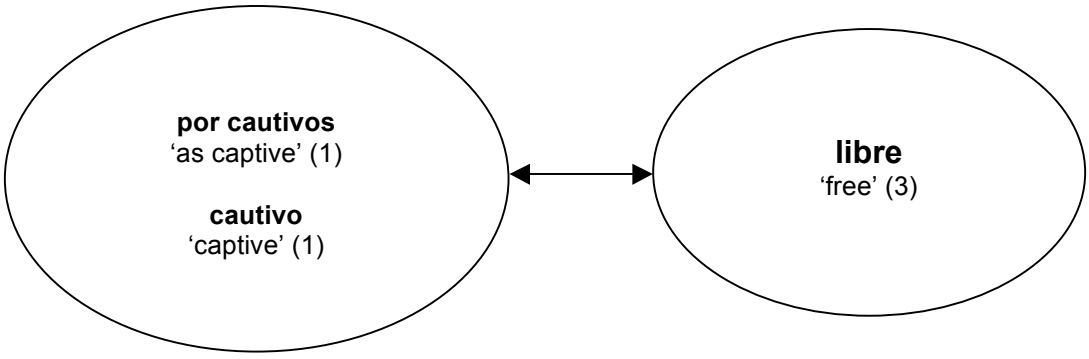


Figure 14, 1300's: The *libre* clusters.

The data from the 1300's show fewer types than those of the 1200's. However, the idea of being made captive is still present and the type *libre* 'free' appears in the data

for the first time. *Libre*, with three occurrences, has a relatively higher frequency and is the central member even though it appears below the proposed CIC threshold of four for this century. It is also the proposed central member in each subsequent century.

Example (63) shows how King Periamo and his city will be free only after he fulfills his part of a bargain negotiated with the Greeks. It is a change of state in which the Greeks are the agents of the change. If he fulfills his obligations, King Periamo and his city will be granted freedom, if not, they will presumably remain under Greek domain.

(63) *Los condes se tornaron a troya al rrey periamo & dyxeronle commo aujan tratado con los griegos que el diese c'iento c'inquenta mjll marcos de oro ...E otrosy a elena & a todas las donzellas que conella fueran tomadas . E con esto que el quedase libre & toda su c'ibdat.*
'The counts returned Troy to King Periamo and they told him how they had negotiated with the Greeks that he would give them one hundred fifty thousand gold marks… And also Helen, and all of the damsels that had been taken with her. And with this he would be free as would all of his city.' (*Sumas de la historia troyana.* 14th c. Leomarte. O'Neill 1999)

5.2.2.3 The *libre* clusters in the 1400's

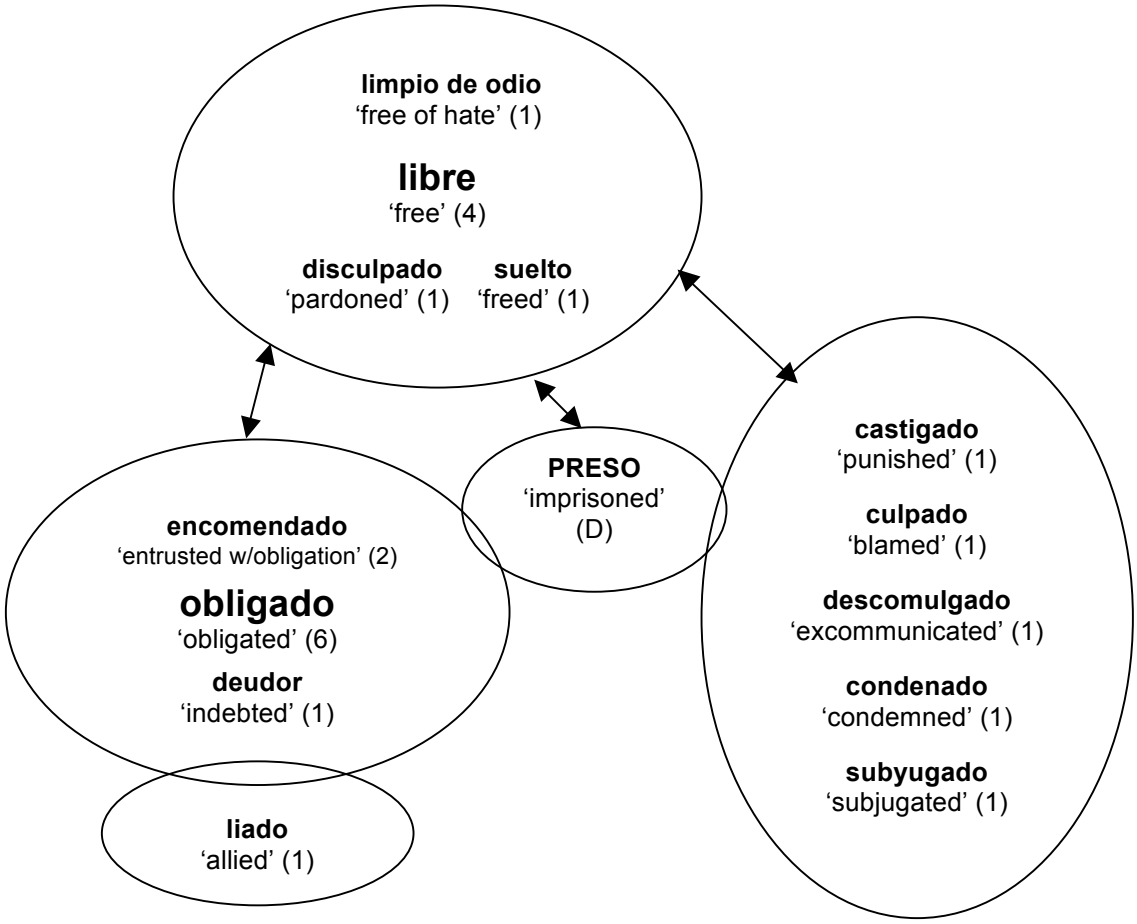


Figure 15, 1400's: The *libre* clusters.

As seen in Figure 15, there is a noticeable expansion of types found it the data from the 1400's. There are two types with token frequency above the CIC threshold of 3 for this century: *libre* 'free' and *obligado* 'obligated'. In the data, the type *obligado* 'obligated' appears for the first time in this century and will persist through the 1600's, disappear in the 1700's, and re-appear in the 1800's. Despite this, *obligado* appears above the CIC threshold only in this century and the next.

As mentioned in Section 5.1, above, there are a variety of factors that are relevant in determining which type is the central member, including frequency, generality, and

endurance. Taking endurance and frequency into consideration, the proposed central member in this century is *libre* despite the fact that it appears with a lower token frequency than *obligado*. The fact that *libre* appeared in the preceding century (*obligado* did not), and that *libre* has a higher standardized overall frequency than *obligado* in the *Corpus del español* in this, and every other century (see Chart 3, p. 151), it is more plausible that *libre* is the central member of this overall category.

It was an unexpected result that *preso* 'imprisoned' was absent from my data in this century. As seen in Chart 3 (p. 151) *preso* occurred in the *Corpus del español* (Davies 2002-) at a higher rate than *obligado* 'obligated' in the 1400's (1.1 occurrences pre million words vs. 0.7 per million respectively). Given that one of the goals of this study is to portray exemplar categories that speakers could have acquired through usage, it is important to note the following: just because a type did not appear in my data does not mean that it wasn't robustly present in usage. Since frequent members are proposed to play an important part in the emergence of categories by enduring through time and serving as models of analogical extension, it is worthwhile to look at other data sets to see what is going on. Because of the fact that *quedar(se) preso* appears with higher standardized frequency than *quedar(se) obligado* in the *Corpus del español* (Davies 2002-) the following practice is adopted here: (a) the type *preso* is included in Figure 15 in order to demonstrate that it was present in other data from this century, and (b) *preso* is in capital letters, and a 'D' is placed between the parentheses in order to indicate that it comes from the *Corpus del español* (Davies 2002-). This practice will be repeated in the 1700's, but with *obligado* and *libre*. Types included in the figures through coordination with the *Corpus del español* will not be included in the analysis of type frequency.

The *libre* cluster contains types that describe states that are related to the subject becoming free from an authority figure. The adjective *perdonado* 'pardoned', as seen in Example (64), describes a state granted on the queen in which she is not only free from a previous accusation leveled by a governing body, but able to serve as an amending force of previous cruelty.

(64) *Queda, luego, no sólo desculpada la reyna que tanto beneficio a sus fijos procuró, mas tan emendadora de la crueza.*
'[She] becomes, then, not only pardoned, the queen that procured so much benefit for her children, but so amending of the cruelty.' (*Las Siete edades del mundo*. 15th c. P. de Santa María. O'Neill 1999)

The adjective *obligado* 'obligated' is proposed to have an opposite value from *libre* because it is a constraint on the subject's freedom. Even though it may not go as far in the opposite direction as to signify physical captivity, it does point to an externally imposed duty to carrying out a task. The subject of Example (65) undergoes a change of state in which he takes on brotherly obligations. This sense of obligation receives additional emphasis by appearing in tandem with the adjective type *deudor* 'indebted' which is defined by the DRAE as: *adj. Que debe, o está obligado a satisfacer una deuda* 'that owes, or is obligated to satisfy a debt'. Here, the agent of the change is the subject and not an authority figure, but he has become beholden to the social norms that dictate concepts of what it means to be obligated to a brother.

(65) *Assí como creo es hecho, antes le quedo deudor e obligado para quanto biua, no como a criado de mi padre, pero como a verdadero hermano.*
'This is how I believe it is done, before I become debtor and obligated to him as long as he lives, not like a servant of my father, but like a true brother.' (*La Celestina*. 15th c. F. de Rojas. BVMC)

In the data the type *liado* 'allied' is connected to the *obligado* 'obligated' cluster because it represents a type of obligation; allies must go to war for one another. The two

kings from Example (66) formalize this obligation by swearing an oath. Through this process they enter a state of alliance.

(66) *Tomaronle juramento en forma personalmente & ansi quedaron liados los dos Reyes.*
'They took the oath in each others presence & this is how the two kings became allied.' (*La Celestina*. 15th c. F. de Rojas. BVMC)

The five types, *castigado* 'punished', *culpado* 'blamed', *condenado* 'condemned', *subygado* 'subjugated', and *descomulgado* 'excommunicated' are proposed to represent opposites to *libre* 'free', but in a different way than *obligado* 'obligated'. All of these adjectives describe resultant states in which certain privileges and / or freedoms have been taken away by an authority, and additionally implicate some sort of consequence or punishment. The subject of Example (67) is imploring the Madame not to impose her authority to blame and condemn. Even though being blamed is not as far along in the process as being condemned is, it can be part of a progression in which someone ends up being persecuted in some way.

(67) *Por Dios, señora, que me dexes concluyr mi dicho, que ni él quedará culpado ni yo condenada.*
'By God, Madame, that you let me conclude what I was saying, that neither he shall be blamed nor I condemned.' (*La Celestina*. 15th c. de Rojas, F. BVMC)

5.2.2.4 The *libre* clusters in the 1500's

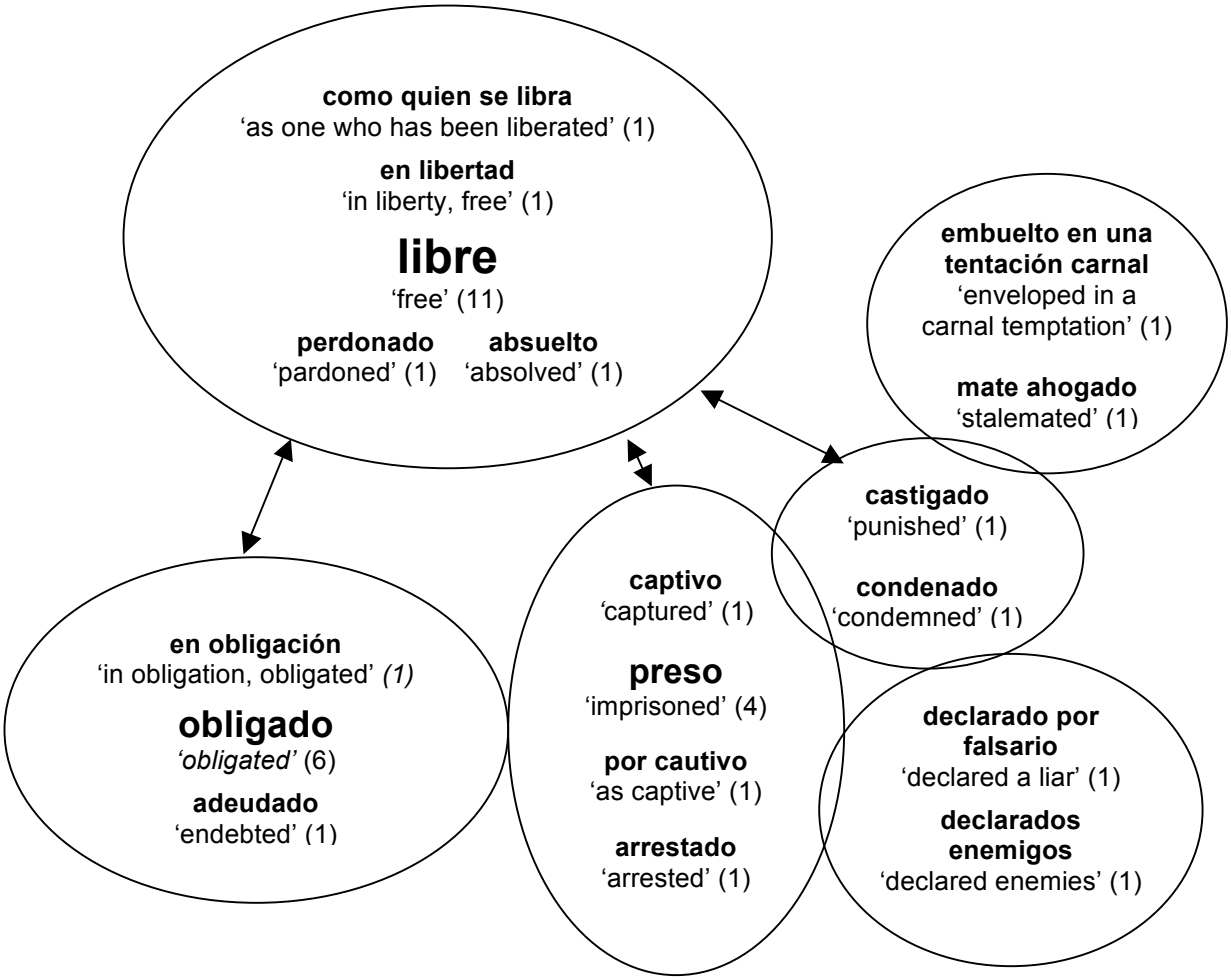


Figure 16, 1500's: The *libre* clusters.

With eleven tokens, *libre* 'free' is proposed to be the central member of this category in this century. There are two other members of the *libre* exemplar clusters in the 1500's that serve as the central members of the clusters in which they appear: *obligado* 'obligated', and *preso* 'imprisoned'. All of these, *libre, obligado*, and *preso*, surpass the proposed CIC threshold for this century of four or more tokens. These CIC's are surrounded by related types, each with only one occurrence, as shown in Figure 16.

Except for the resurgence of *preso* 'imprisoned' and related types, this set of clusters highly resembles the *libre* 'free' clusters from the previous century in its structure. Types indicating punishment persist (e.g. *condenado* 'condemned', *castigado* 'punished'), and both *libre* 'free' and *obligado* 'obligated' appear in their own clusters surrounded by related types.

Becoming free is not just a change of state that is granted by an outside authority with respect to the prevailing laws. For the first time in these data, *quedar(se) libre* 'to become free' is used to describe a change of state in which the subject becomes free of a situation, in this case emotional, without having an authority figure act as an intermediary of the change. Example (68) portrays a subject that becomes free of her distress by having a child.

(68) *En esto, plugo a Dios que, invocando el nombre de nuestra señora, parió un muy hermoso hijo, y quedó libre de la congoxa en que estava, e muy alegre con el fijo nacido.*
'Because of this, she begged God, invoking the name of Our Lady, she gave birth to a beautiful son, and became free of the distress that she was in, and very happy with the son that was born.' (*Laberinto de amor*. 16th c. (1546). J. Boccaccio. Lemir)

In Example (69), the idea of freedom does not have to do with the subject's legal status. Although brought about by the king, it is the absence of his apparently oppressive person, and not a legal ruling, that leaves Osita feeling like someone freed from a shipwreck.

(69) *Dexando a Osita [el rey] se fue con sus caçadores a buscar el ciervo, donde se detuvo algunos días. Quedó la santa donzella como quien se libra de un terrible naufragio.*
'Leaving Osita [the king] left with his hunters to look for the stag, where he stayed a few days. The good damsel became like someone who has been freed from a terrible shipwreck.' (*Fructus Sanctorum y Quinta Parte del Flos Sanctorum*. 16th c. (1594). A. de Villegas. Lemir)

The idea of being a prisoner also has more metaphorical connotations. The subject of Example (70) views himself as a prisoner of the woman that he loves. Although his being captive is metaphorical and not legal, the comparison creates a situation in which the woman indicated possesses symbolic authority.

(70) *por vuestra voluntad vivo, vuestro querer me sustenta y mi vida pende de vuestras manos. Las quales, humildemente besando, quedo por vuestro cautivo.*
'By your will I live, your love sustains me and my life hangs from your hands. Those which, humbly kissing, I am left as your captive.' (*Comedia llamada Florinea*. 16th c. (1554). J. Rodríguez Florián. Lemir)

The types *castigado* 'punished' and *condenado* 'condemned' were placed in a separate but intersecting cluster to *preso* 'imprisoned' because they do not specifically refer to being imprisoned. The action of being condemned or punished could imply a variety of punitive actions. A person could be condemned to death, banishment, or to imprisonment. Punishment has many forms as well, some of which may (or may not) include imprisonment. The hypothetical judge in Example (71) will be condemned if he takes the implicated party's blame for granted. Yet, there is no indication as to what fate the judge will be condemned. Because it is a proverb, and proverbs expound common wisdom or values, the condemnation would come from a hypothetical society as perceived by Séneca.

(71) *Dezía Séneca en sus Proverbios que el juez quedava condenado cuando dava por libre al culpado.*
'Séneca said in his Proverbs that the judge became condemned when he treated the blamed person as free.' (*Fructus Sanctorum y Quinta Parte del Flos Sanctorum*. 16th c. (1594). A. de Villegas. Lemir)

The cluster containing *declarado por falsario* 'declared a liar' and *declarados enemigos* is placed in a position that relates to being imprisoned, punished, and

condemned. It is possible that part of being condemned, or punished, is that the offender is also declared a liar or declared an enemy. In example (72), the accusers are not only left without fame, without life, and left under a pile of rocks, but they are declared to be liars as a final condemnation.

(72) *Quedó con fama$_1$ y con vida$_2$, y sus acusadores, sin vida$_3$ y sin fama$_4$, cubiertos de piedras$_5$ y declarados por falsarios$_6$.*
'He came out with fame$_1$ and with life$_2$, and his accusers, without life$_3$ and without fame$_4$, covered in rocks$_5$, and declared as liars$_6$.' (*Fructus Sanctorum y Quinta Parte del Flos Sanctorum*. 16th c. (1594). A. de Villegas. Lemir)

Located in close proximity to the *preso* 'imprisoned' cluster and the *castigado / condenado* 'punished / condemned' cluster, the types in the cluster with *embuelto* 'enveloped' and *mate ahogado* 'stalemated' convey an idea of being trapped metaphorically, but not legally, as in Example (73). By becoming enveloped in temptation, God's servant finds himself trapped by an uncommon phenomenon.

(73) *El siervo de Dios quedó embuelto en una tentación carnal tan vehemente y furiosa que en su vida tuvo otra semejante.*
'God's servant became enveloped in a carnal temptation so vehement and furious that never in his life did he have a similar one.' (*Fructus Sanctorum y Quinta Parte del Flos Sanctorum*. 16th c. (1594). A. de Villegas. Lemir)

The examples presented from this century depict changes of state in which the subject gains or loses freedom in a metaphorical sense. This is intended to demonstrate to the reader that the usage of adjective types in *quedar(se)* + ADJ is not restricted to only certain semantic domains that a particular adjective may have. In other words, as *libre* emerges as an established adjective type, all of the implications of this adjective can be seen in usage. A person can be freed from prison, freed from a net, freed from an

emotional situation (Ex. (68)), or metaphorically free (i.e. the simile depicted by the verbal usage of *librarse* 'to free one's self' in Example (69)).

5.2.2.5 The *libre* 'free' clusters in the 1600's

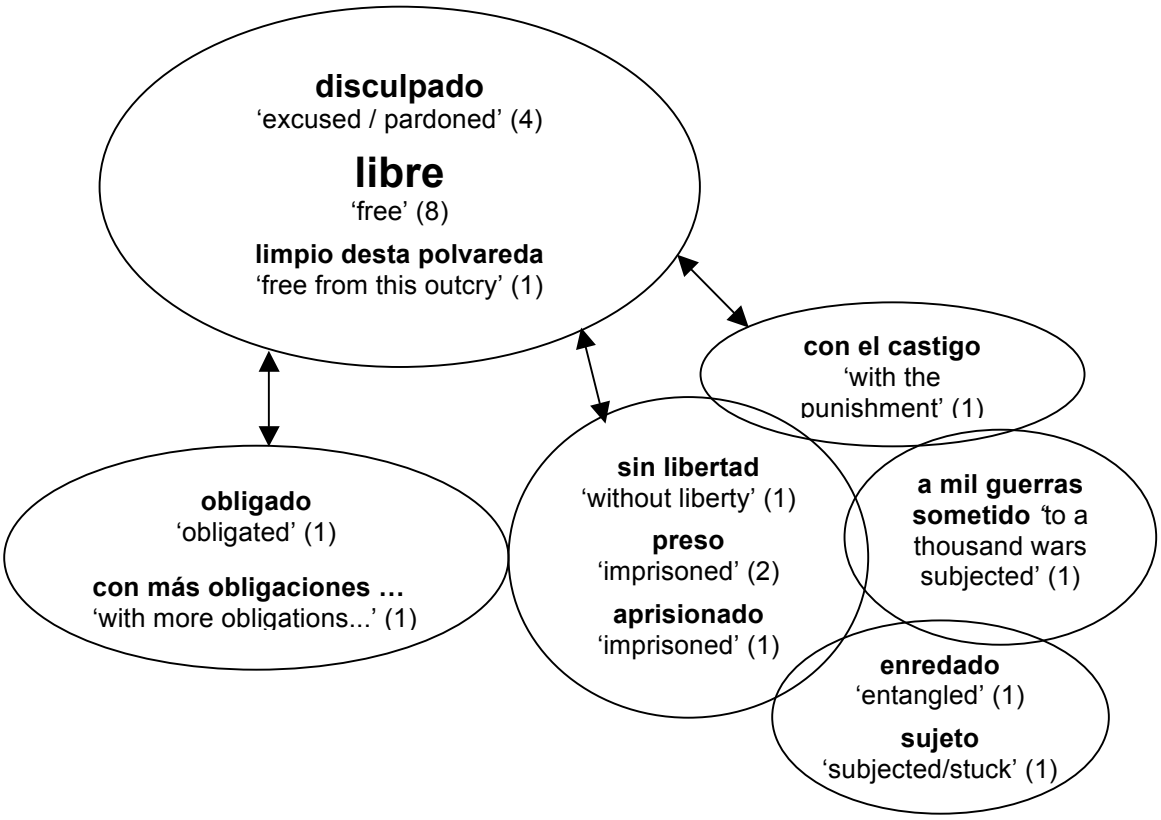


Figure 17, 1600's: The *libre* clusters.

In its overall arrangement of adjective types, the proposed clusters from the 1600's resemble the ones from the previous centuries. However, neither *preso* 'imprisoned' or *obligado* 'obligated' surface as clear central members of their particular clusters in this century. *Libre*, with eight occurrences in the data, is clearly the central member of this category and is the only type, except for *disculpado* 'excused / pardoned', that appears above the CIC threshold of three for this century.

As in the previous century, *quedar(se) libre* can refer to becoming free from an authority figure, as in Example (74), or from an emotional situation, as in Example (12).

(74) *Como siendo él quien logre el triunfo más alto hoy en mi servicio, quedo libre.*
'As being he who achieved the highest triumph today in my service, he is free.' (*Afectos de odio y amor*. 17th c. P.C. de la Barca. BVMC)

(75) *Ambas graciosas son, ambas son bellas.*
De verme Amor se ríe y, mientras, temo
que aguda flecha en mis entrañas vibre.

Aunque tengo delante dos estrellas,
sin norte voy y, en fin, en tal extremo,
no sabiendo qué hazer, me quedo libre.

'Both are graceful, both are beautiful.
Love laughs upon seeing me and, meanwhile, I fear
that a sharp arrow quivers in my entrails

Even though I have before me two stars,
without a guiding star I go and, finally, in such an extreme,
not knowing what to do, I set myself free.' (*La Constante Amarilis*. 17th c. (1609). C. Suárez de Figueroa. Lemir)

The types *enredado* 'entangled' and *sugeto* 'subjected / stuck' come from a single example (76) that describes how love may leave the lovers entangled, hurt, and subjected to (or stuck in) a situation in which they will never be satisfied. These types are in a separate cluster because they are metaphors for losing freedom without actually indicating that the subjects were imprisoned or punished.

(76) *Ved qué embuste de Amor, que desta suerte*
quedaron ellas y ellos enredados,
heridos de la flecha dura y fuerte,
y sugetos a no ser bien pagados.
'See what a lie of love, that this way
they became entangled,
wounded by the arrow hard and strong,
and subjected to never being satisfied.' (*Genealogía de la toledana discreta*. 17th c. E. Martínez. Lemir)

5.2.2.6 The *libre* 'free' clusters in the 1700's

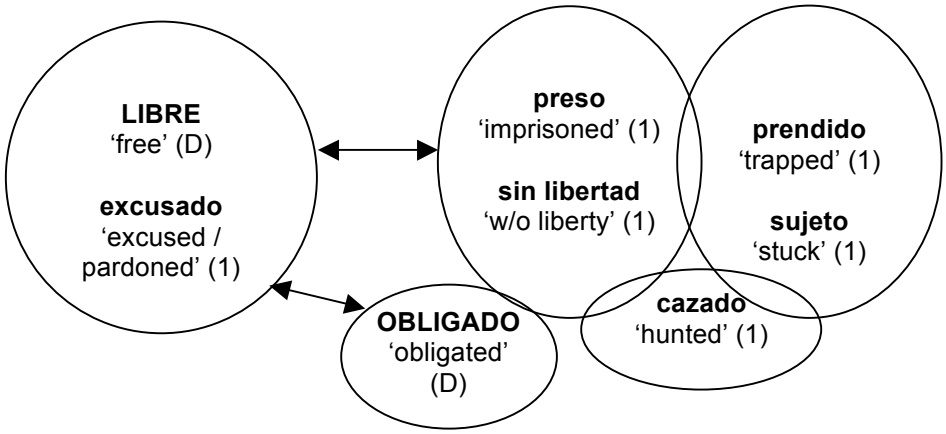


Figure 18, 1700's: The *libre* clusters

As alluded to in the introduction to this section on the *libre* clusters (5.2.2), the 1700's presents an idiosyncratic case for the *libre* clusters. *Quedar(se) libre* 'to be set free' does not appear in my data despite the fact that it occurs in surrounding centuries at well above the proposed CIC threshold (8 total occurrences in the 1600's, 7 in the 1800's). Using the rationale and notation described in Section 5.2.2.3, the type *quedar(se) libre* is included in these clusters because of it's robust standardized frequency in the *Corpus del español* (Davies 2002-: 8.5 occurrences pre million words). Considering the relatively high frequency of *libre* in surrounding centuries, and the aforementioned standardized frequency, it is proposed here that *libre* continues to be the central member of this category in this century. Similarly, *quedar(se) obligado* 'to be obligated' did not appear in my data, but was included because of its 3.4 occurrences per million words in Davies' (2002-) corpus.

The types *sujeto* 'stuck' and *prendido* 'trapped' in this century have to do with the subject becoming trapped (or stuck) in an undesirable situation, as in (77) and (78). Because the type *sujeta a una triste viudez* 'stuck in a sad widowhood', from Ex. (77),

also implies entering a state of widowhood, a state in which the subject is left without their spouse, this type is also included in the *solo* clusters (Figure 11, p. 118). This demonstrates the links between the different clusters discussed above (Section 5.1).

(77) *...y después que, asegurados todos de la verdad del caso por muchas verídicas cartas y no menos verdaderos testigos, quedó mi desconsolada madre sujeta a una triste viudez.*
'…and after which, all assured of the truth of the matter by true letters and no less truthful witnesses, my disconsolate mother became stuck in a sad widowhood.' (*Los trabajos de Narciso y Filomela*. 18th c. V. Martínez Colomer. BVMC)

(78) *... y al momento comencé a prevenir las redes en que yo misma quedé prendida.*
'…and immediately I began to prevent the tangles in which I myself became trapped.' (*Los trabajos de Narciso y Filomela*. 18th c. V. Martínez Colomer. BVMC)

The type *excusado* 'excused / pardoned' describes an outcome in which the person in question would be freed from imprisonment and punishment, as in Example (79).

(79) *y verás que nos hallamos*
en casa con dos docenas
de ministros y criados,
y le encajan en la cárcel
atado de pies y manos.
LESMES: Tío infernal, ¿no habrá un medio
para quedar escusado?
'and you will se that we shall find ourselves
at home with two dozen,
ministers and servants,
and they will put you in jail
with hands and feet tied.
Lesmes: Infernal Uncle, isn't there a way
to be pardoned?' (*El celoso don Lesmes: comedia nueva en tres actos*. 18th c. D.Vicente Rodríguez de Arellano. BVMC.)

5.2.2.7 The *libre* 'free' clusters in the 1800's

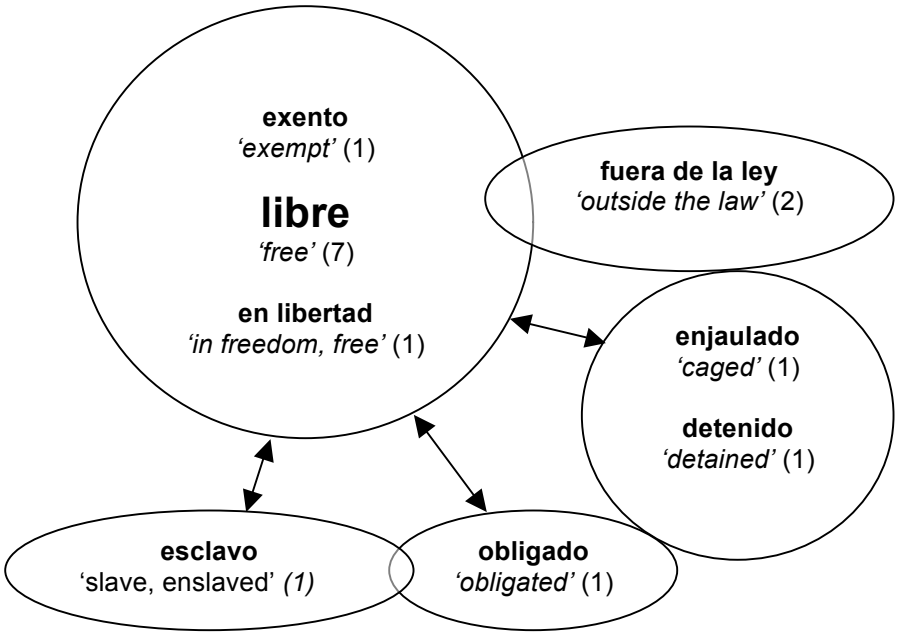


Figure 19, 1800's: The *libre* clusters.

The *libre* clusters in the 1800's show a similar structure to the same clusters from the 1400's to the 1700's. *Quedar(se) libre* 'to be set free' is clearly the central member of this category and, with seven tokens, appears above the proposed CIC threshold of four for this century. As seen in Figure 19, the central *libre* cluster is surrounded by types that are included in this set of clusters by virtue of being opposites of *libre*.

They type *fuera de la ley* 'outside the law' presents an interesting case because it places the subjects somewhere between freedom and imprisonment (and/or punishment). As seen in Example (80), the subjects are not in a state of imprisonment, nor are they totally free; this type is proposed to relate to both the *libre* and its opposites that indicate a loss of freedom. The Municipality, the General, and all who help them are to be considered outside the law and it is assumed that the authorities will attempt to bring them to justice.

(80) *Los miembros de la Municipalidad de París, el comandante general Henriot y todos los que favorezcan la insurrección o la auxilien en sus proyectos liberticidas quedan igualmente fuera de la ley.*
The members of the Paris Municipality, the Commander General Henriot, and all who favor the insurrection or help in their liberty destroying projects are equally outside the law.' (*Amor de padre*. 19th c. F. Martínez. BVMC)

Becoming enslaved represents a loss of freedom in which the subjects are obligated to carry out the master's will. Example (81) shows how vulnerable populations in a war torn region are at risk of becoming enslaved.

(81) *Fácil cosa es pelear... pero cuando se ve uno rodeado de poblaciones enteras, sin abrigo ni amparo, extenuadas de cansancio y de hambre; cuando no se ven por todas partes sino mujeres y niños demandando socorro a gritos, y amenazados de quedar esclavos.*
It's an easy thing to fight... but when one finds himself surrounded by entire populations, without coat or protection, extenuated by exhaustion and by hunger; when women and children are seen all over the place screaming for help, and threatened by becoming enslaved.' (*Aben Humeya o La rebelión de los moriscos*. 19th c. D. Francisco Martínez de la Rosa. BVMC)

5.2.2.8 Overview of the *libre* 'free' clusters

Over time, these clusters emerged as a set of clusters with a clear central member, *quedar(se) libre* 'to be set free', that appeared in my data above the CIC threshold from the 1400's to the 1800's (with the exception of the 1700's). It was also above the prefab threshold from the 1300's to the 1800's in Davies' (2002-) data. Although the prefab *quedar(se) libre* shows endurance, it does not show a steady gain in standardized frequency as *quedar(se) solo* 'to be left alone' was demonstrated to do in Section 5.2.1.[23] Instead, *quedar(se) libre* emerged in the data, gained in frequency, and then waned, as will be discussed below. Nevertheless, the fact that *quedar(se) libre* has appeared above

[23] *Quedar(se) solo* went from occurring 0.04 times per 10k words in the 1200's, to a peak of occurring at 0.30 times per 10k words in the 1800's.

the prefab threshold in Davies' (2002-) data over a span of six centuries provides another case to support the proposition that prefabs have longevity.

In order to examine the prefab status of *quedar(se) libre* Table 16 shows the occurrences per million words of the adjective *libre* 'alone' in the first column, the percentage of all of these occurrences of *libre* accounted for by *quedar(se) libre* in the second column, the occurrences per million words of *quedar(se)* + ADJ in the third column, and the percentage of *quedar(se)* + ADJ accounted for by *quedar(se) libre* in the fourth column which also provides the occurrences per million. The final column indicates the centuries in which *quedar(se) libre* appeared above the prefab threshold. All figures come from the *Corpus del español* (Davies 2002-). Only in two centuries, the 1300's and the 1900's, did *quedar(se) libre* fall short of the prefab threshold of accounting for 1% of all occurrences of *libre*, and 4% of all occurrences of *quedar(se)* + ADJ. The peak for *quedar(se) libre* appears to occur in the 1700's in which it accounts for 4.13% of all occurrences of the adjective *libre* and 8.6% of all occurrences of the construction *quedar(se)* + ADJ.

Table 16. Figures used in calculating prefab threshold of *quedar(se) libre* (Davies 2002-)

	# *libre* per M	*quedar(se) libre* (%) / *libre*	# *quedar(se)* + ADJ per M	# *quedar(se) libre* (%) / *quedar(se)* + ADJ	Prefab status
1200's	213.4	0.05%	3.0	0.1 (3.3%)	
1300's	94.4	1.17%	25.5	1.1 (4.3%)	x
1400's	141.2	1.13%	35.3	1.6 (4.5%)	x
1500's	229.1	4%	109.2	9.2 (8.4%)	x
1600's	242.1	4.13%	115.8	10 (8.6%)	x
1700's	265.8	3.2%	111.2	8.5 (7.6%)	x
1800's	267.7	1.18%	108.4	4.8 (4.4%)	x

x Surpassed the prefab threshold

In the preceding section presenting the *solo* clusters, it appeared that *quedar(se) solo* was a CIC before becoming a prefab. This is logical since the formula for identifying prefabs is more exclusive. However, here are some discrepancies between the two frequency measurements regarding *quedar(se) libre*, as will be seen below in Table 19. It appears that it achieves prefab status before it becomes a CIC; it is a prefab in the 1300's, but not a CIC until the 1400's. This is due to the fact that the two measurements of frequency are applied to two different sets of data, the CIC measurement to my own data, and the prefab measurement to Davies' (20002-).

Looking at token frequency, Table 17 shows that, despite its longevity, it appears that *quedar(se) libre* 'to be set free' experienced a peak in the 1500's and then began to decline. This table provides the total number of tokens per century in the first column, the number and percentage of the total tokens that belong to *quedar(se) libre* in the second column, the overall frequency of *quedar(se)* + ADJ in my data as occurring per 10k words in the third column, and the overall frequency of *quedar(se) libre* in my data as occurring per 10k words in the fourth column. The standardized frequency figures in the fifth column for the adjective *libre* come from Davies' (2002-) data.

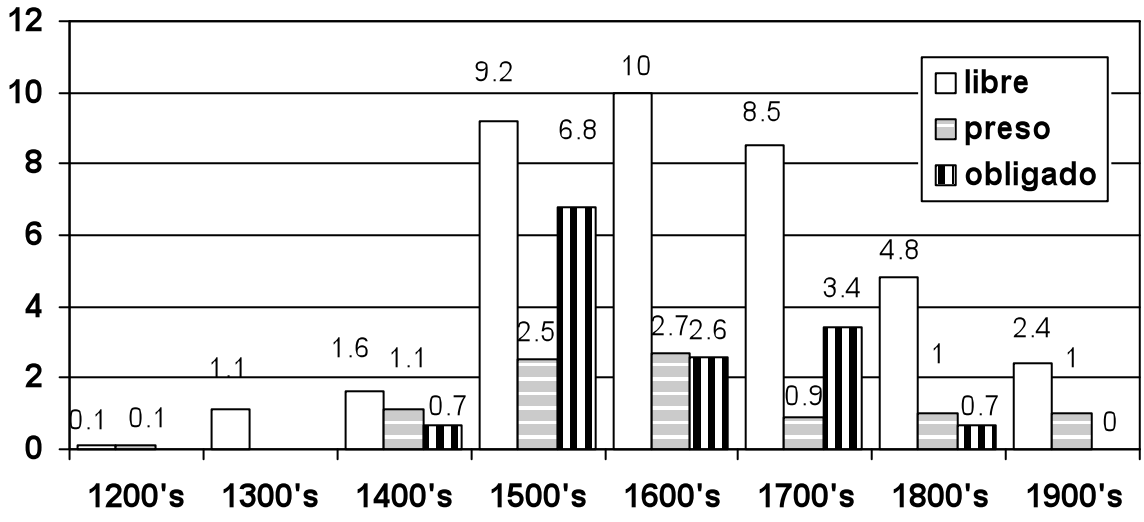
Table 17. Number and percentage of tokens of *quedar(se) libre* per century.

	Total # of tokens	# of tokens *quedar(se) libre* (% total N of tokens)	# *quedar(se)* + ADJ per 10k words	# *quedar(se) libre* per 10k words	# *libre* per 10k words
1200's	43	---	0.42	---	2.13
1300's	76	2 (2.5%)	0.96	0.03	0.94
1400's	196	4 (2%)	1.18	0.02	1.41
1500's	271	11 (4%)	4.85	0.20	2.3
1600's	280	8 (3%)	6.00	0.17	2.42
1700's	253	---	5.82	---	2.66
1800's	255	7 (2.5%)	2.64	0.07	2.68
1900's*	181	---	1.83	---	2.1

* From Bybee & Eddington (2006), except for last column.

Unlike *quedar(se) solo*, which showed a steady rise over the centuries studied, *quedar(se) libre* goes from occurring at a very low overall frequency in the 1300's and 1400's (0.03 and 0.02 times per 10k words, respectively), to almost ten times that rate in the 1500's. It appears that the 1500's represents a peak, as far as overall frequency, before *quedar(se) libre* begins to decline, occurring at a lower rate in subsequent centuries. By the 1800's, it occurs at a rate of only 0.07 per 10k words. In terms of the percentage of tokens *quedar(se) libre* accounts for, there is a similar peak. In the 1500's, it accounts for 4% of all tokens of *quedar(se)* + ADJ while in surrounding centuries it accounts for noticeably less. In the 1900's, from Bybee & Eddington (2006), there were no tokens of *quedar(se) libre* in their data, nor did it occur with any other adjectives they studied. These two peaks in frequency were also visible in Davies' (2002-) data (see Table 16 and Chart 3)

Chart 3. Standardized frequency of *quedar(se)* plus *libre*, *preso*, and *obligado* in 8 centuries of the *Corpus del español* (Davies 2002-); occurrences per 1,000,000 words.



Checking these findings of frequency against the *Corpus del español* (Davies 2002-), as shown in Chart 3, there is similar evidence of a rise and fall in the overall frequency of *quedar(se) libre* 'to be set free'. Furthermore, Chart 3 includes the types *quedar(se) obligado* 'to be obligated' and *quedar(se) preso* 'to become imprisoned', two key types in the *libre* clusters that showed endurance over time[24]. The figures for the 1900's come only from the fictional writing.

As with my own data, there is a leap in the rate of standardized frequency of *quedar(se) libre* from the 1400's to the 1500's. However, in the *Corpus*, the overall frequency peaks in the 1600's, instead of the 1500's, before declining to 2.4 occurrences per million words in the 1900's. This search shows that my own data for 1700's are indeed idiosyncratic and that a more likely scenario is that *quedar(se) libre* fit into a larger pattern of a steep increase in frequency, peaking in the 1500's or 1600's, followed by a steady decrease in subsequent centuries.

Quedar(se) preso occurred only one time in Davies' (2002-) data in the 1200's, giving it a standardized overall frequency of 0.1 occurrences per million words. Neither *preso* nor *obligado* appeared with *quedar(se)* in the 1300's. The peak centuries for both of these types coincide with the two most robust centuries for *quedar(se) libre*: the 1500's for *quedar(se) obligado* (6.8 occurrences per million words) and the 1600's for *quedar(se) preso* (2.7 per million). These two centuries represent the massive rise and peak in standardized frequency for *quedar(se)* + ADJ. Both of these types drop in frequency from their peak centuries to the 1900's, in which *quedar(se) preso* appears at a

[24] A search was done for all occurrences of the verb *quedar* appearing with *libre*, *obligado*, and *preso* both plural and singular. A portion of these pairings was individually analyzed in order to verify that a majority of these combinations conveyed a change of state with a human subject.

low frequency (1 per million) and *quedar(se) obligado* disappears from Davies' (2002-) data.

Table 18. Types in the *libre* clusters

	Total # types	Types in *libre* clusters
1200's	29	5 (17%)
1300's	40	3 (7.5%)
1400's	121	13 (10.5%)
1500's	153	16 (10.5%)
1600's	171	12 (7%)
1700's	135	6 (4.5%)
1800's	141	8 (5.5%)
1900's*	54	--

* From Bybee & Eddington (2006)

In terms of type frequency, Table 18 shows the total number of adjective types occurring in *quedar(se)* + ADJ in my own data per century in the first column, with the number of types proposed to belong to the *libre* clusters, and the percentage of all tokens that these represent in the second column. Even though types attributed to the *libre* clusters account for 17% of all types in the 1200's, they drop down to accounting for 7.5% of all types in the 1300's. After the type frequency of this category increases to represent a total of 10.5% of all types in the 1400's and 1500's, it tapers off in these data. By the 1800's, types proposed to pertain to the *libre* clusters account for only 5.5% of all types.

As with the *solo* clusters, it appears that there is a correlation between the token frequency of the central member of the category and the type frequency of that category. In this case, however, the rise in type frequency precedes the conspicuous rise in the overall frequency of the central member *quedar(se) libre* 'to be set free'. Regarding token frequency of *quedar(se) libre*, it goes from 0.02 occurrences per 10k words in the 1400's

to 0.20 in the 1500's in my data (see Table 17). The types in the *libre* clusters go from accounting for 7.5% of all types in the 1300's to accounting for 10.5% of all types found in the data in the 1400's and 1500's.

Other types included in the *libre* clusters show longevity in the data as well. Table 19 shows types from these clusters that occurred in three or more centuries in these data. *Quedar(se) obligado* 'to be obligated' appears in the data from the 1400's to the 1800's (excepting the idiosyncratic 1700's in which it appeared only in the *Corpus* (Davies 2002-)). In the first two centuries in which it appears, the 1400's and the 1500's, it does so above the CIC threshold. After that, it appears in the data with a lower token frequency. Even though *quedar(se) preso* 'to become imprisoned' first appears in the 1200's, it appears neither in these data nor in Davies' (2002-) in the 1300's. Although it appears in Davies' data in the 1400's, it resurfaces in my own data above the CIC threshold in the 1500's and continues to appear in the 1600's and 1700's albeit below the CIC threshold. The type *por cautivo* 'as captive' does not appear above the CIC threshold in these data, but still appears over a considerable span of time. None of these types appeared in Bybee & Eddington (2006) with any of the four verbs they studied.

Table 19. Types from the *libre* clusters that appear in at least 3 centuries.

	1200	1300	1400	1500	1600	1700	1800	1900
libre 'free'	DDD	====	####	####	####	DDD	####	DDD
obligado 'obligated'			####	####	====	DDD	====	DDD
preso 'imprisoned'	====		DDD	####	====	====	DDD	DDD
por cautivo 'as captive'	====	====		====				

(====) = presence of type in data below CIC threshold
(####) = presence of type in data above CIC threshold
(DDD) = presence only in Davies *Corpus del español* (2002-)

Once again, the *libre* clusters demonstrate that formulaic linguistic sequences and the categories to which they belong endure over great spans of time. However, unlike the

solo clusters, this category shows evidence not only of expansion but of contraction. The central member *quedar(se) libre* emerges in my data in the 1300's and in Davies' (2002-) in the 1200's and is the proposed central member through the span of the data. There is a sharp increase in the standardized frequency of *quedar(se) libre* from the 1400's to the 1500's (see Table 17 and Chart 3). These two centuries are also the centuries in which there is a rise in the type frequency of the *libre* clusters (see Table 18), providing evidence that there is a correlation between the frequency of the central member and type frequency of the category. This is part of a larger trend in which *quedar(se) libre* gains in standardized frequency, peaks in the 1600's, and begins to wane. It appears that there is also a correlation between the drop in the frequency of the central member and the contraction of the category to which it pertains. Even though it is not a steady drop (there is a slight rise in type frequency in the 1800's, see Table 18), as the standardized frequency of *quedar(se) libre* decreases, so does the type frequency of the category. Other key types, *quedar(se)preso* and *quedar(se) obligado*, diminish in standardized frequency in this phase with the latter disappearing by the 1900's in Davies' (2002-) data (see Chart 3).

5.2.3 Changes of possession: the *rico / pobre* 'rich / poor' clusters

The analysis in this section focuses on clusters centered on two opposite central members: *quedar(se) pobre* 'to become poor', and *quedar(se) rico* 'to become rich'. Viewed this way, this section represents an analysis of two separate sets of clusters that are related through their opposition. Even though opposites have been previously considered, this analysis is different because there is a clear spectrum, and therefore connection, between the two opposites. Whereas *libre* 'free' and *solo* 'alone' do not have

one clear antonym in the data, *rico* 'rich' and *pobre* 'poor' are themselves antonyms. For example, opposites of *libre* could be *preso* 'imprisoned', *obligado* 'obligated', or *con hijo* 'with a child' (as was shown to be the case in the similarity study presented in Chapter 6). This could be due to the fact that *solo* and *libre* are more general than *rico* and *pobre* and, therefore, participate in oppositional relationships with adjectives in a wider variety of contexts.

Even though the idea of changes of possession imply a change in physical attributes, in this set of clusters *quedar(se)* + ADJ is argued to denote a change of social state. As will be demonstrated, almost all of the changes in regards to physical objects or wealth point to changes in social state through the evaluation of a person based on their possessions. When referring to a specific physical object in this construction, it is often a metaphorical change that *quedar(se)* + ADJ implies. For example *quedar(se) sin pan* 'to be left without bread' and *quedar(se) sin camisa* 'to be left without a shirt' (a 'multiple' from Example (95) on p. 171) are proposed to be metaphors for being poor; they do not describe a change in which a certain person literally lost their bread and their shirt (although that possibility is certainly not excluded). Not all changes described by *quedar(se)* + ADJ are metaphorical, however, and some point to a specific object being gained or lost as a result of the change. For the sake of convenience in the discussion, these clusters will be referred to as the *rico / pobre* clusters.

5.2.3.1 The *rico / pobre* clusters in the 1200's

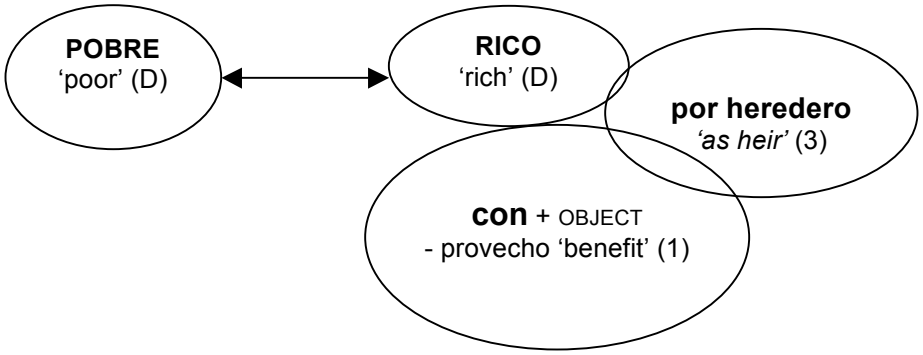


Figure 20, 1200's: Preliminary *pobre / rico* clusters.

As seen in Figure 20, neither *pobre* nor *rico* appeared with *quedar(se)* + ADJ in the data in the 1200's. However, in the Davies (2002-) corpus, there is one occurrence of each in this century. Since this is an indication that the category is in a process of emergence, these are included in these clusters. As was practiced in the previous century, beginning with *preso* 'imprisoned' in the 1500's, *rico* and *pobre* are given in capital letters and the (D) indicates that they come from the *Corpus del español* (Davies 2002-).

The type *quedar(se) por heredero* 'to come out as heir' is another example of a prepositional phrase being used in the adjectival slot of the change-of-state construction *quedar(se)* + ADJ. Types of another possible generalization, *quedar(se) por* + NOUN, are only included in the possession cluster when they denote a change of status that specifically implies wealth, such as *heredera* 'heiress' in Example (82). Whereas a military boss should be rich and have lots of prestige, wealth is not the primary defining factor of whether or not a person is a military boss. However, by becoming heiress, Doña María should necessarily inherit something, most likely great wealth and power. The *heredero* type is also discussed in 5.2.3.3.

(82) '*& a pocos dias murio la reyna & quedo por heredera del reyno do[ñ]a maria la primera hija que fuera hija del marques conrado & quedo don Juan de ybelin por adelantado'*
'And after a few days the queen died, and Doña Maria, the first daughter that was a daughter of the Marques Conrado, was left as heiress of the kingdom, and Don Juan de Ybelin was left as military boss' (*Gran Conquista de Ultramar*. 13th c. Anon. O'Neill 1999)

5.2.3.2 The *rico / pobre* clusters in the 1300's

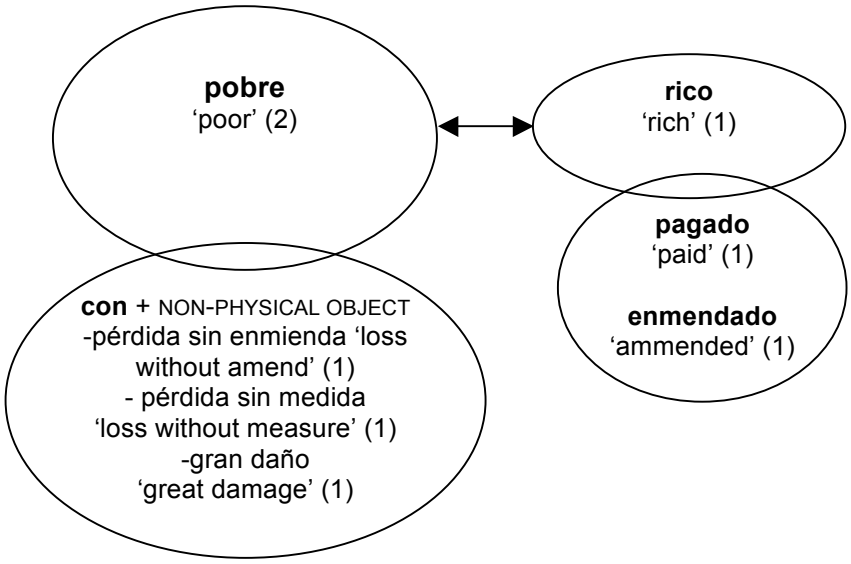


Figure 21, 1300's: The *rico / pobre* clusters.

For the first time in these data, the types *pobre* 'poor' and *rico* 'rich' appear in *quedar(se)* + ADJ. These types provide a clear example of how seeming opposites are conceptually related. Furthermore, they are more general than many other types in these clusters. Example (83) shows how one author compares being poor to being dead.

(83) *'non dexo a su fijo cosa alguna / E tan pobre quedo commo que muerto fuese'*
'(He) didn't leave his son anything / And he was left so poor that it was as though he were dead' (*Sumas de la historia troyana*. 14th c. Leomarte. O'Neill 1999)

An example that illustrates the close relationship that opposites have comes from two of tokens of *quedar(se) con* + NON-PHYSICAL OBJECT in this century: *quedar(se) con*

pérdida sin enmienda ‘to be left with a loss without amend’ and *quedar(se) con pérdida sin medida* ‘to be left with a loss without measure’. Examples (84) and (85) are variations of what appear to be the same proverb given by the same author in different parts of the text to teach the reader a lesson. In both of these, the subject could hypothetically undergo a change of state that would leave him or her with a great, although unspecified, loss. Even though the preposition *sin* ‘without’ would typically point to the subject losing something in *quedar(se)* + ADJ, examples (84) and (85) show how the opposite, *con* ‘with’, can illustrate material loss in this construction.

(84) *mas vale quedar con poca verguenc'a que con perdida syn medida*
‘It’s better to be left with a little shame than with loss without measure’ (*Sumas de la historia troyana*. 14th c. Leomarte. O’Neill 1999)

(85) *E mas vale quedar los omnes algunt poco envergonc'ados que con perdida syn emjenda*
‘And it is better that the men are left a little bit ashamed than with loss without measure’ (*Sumas de la historia troyana*. 14th c. Leomarte. O’Neill 1999)

The type *quedar(se) con gran daño* refers to a situation in which a military body suffers losses, some of which must be material, and is considered similar to the types shown in examples (84) and (85) above. The other types in these examples portray a situation in which the subject is ashamed; these are different from the other types that hint at a conceptual material loss.

Referring back to Figure 21, the types *enmendado* ‘amended’ and *pagado* ‘paid’ are included in this set of clusters because they refers to a change in which the subject is compensated for some type of debt or injustice. Like getting rich, this could be an opposite to ‘being left with a loss without measure’ and is proposed to be related to *rico* ‘rich’ by virtue of being on the end of the scale in which the subject gains something.

5.2.3.3 The *rico / pobre* clusters in the 1400's.

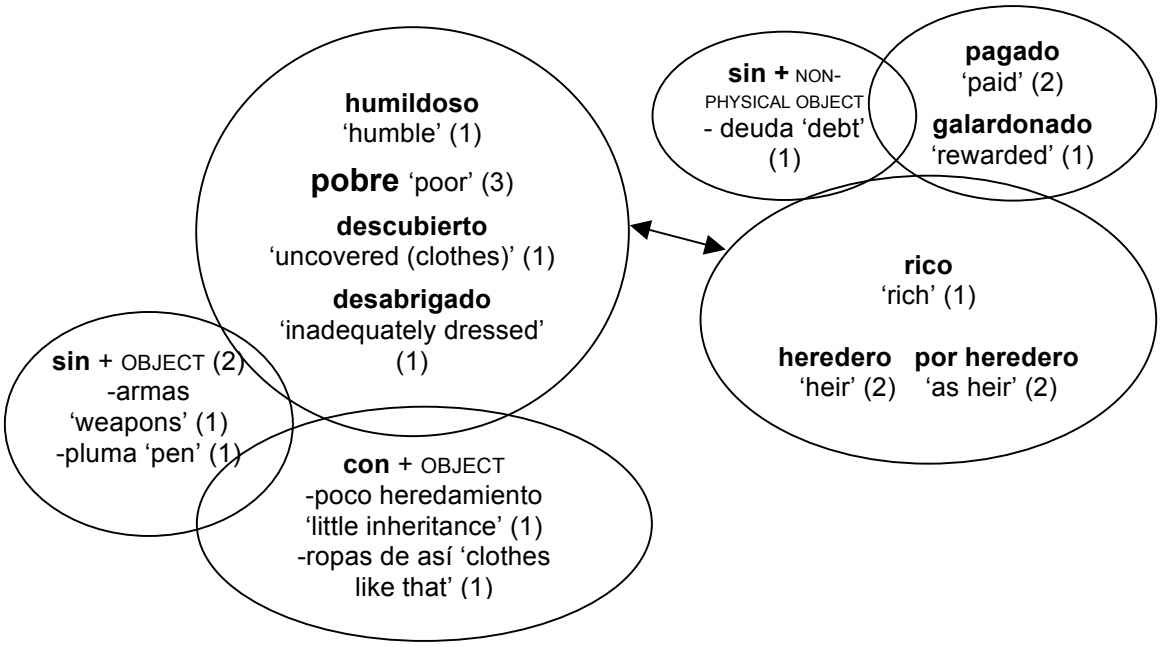


Figure 22, 1400's: The *rico / pobre* clusters.

As in the 1300's, there are clusters of types organized around *pobre* 'poor' and around *rico* 'rich'. At the same time, there is a general category expansion in the data with many new types appearing in the clusters.

All the types included in the *rico* 'rich' cluster reflect a change in which the subject will presumably come out wealthier. Looking at token frequency alone, it is possible that the central member should be *heredero* 'heir'[25] or *por heredero* 'as heir' since they each have two tokens. However, because there is no single member with a conspicuously high token frequency, the low number of tokens in general opens the possibility of considering other criteria. Given that one of the salient characteristics of a central category member is generality, this supports *rico* as the central member. Whereas one may become rich through a variety of paths, becoming heir is more specific because

[25] According to the DRAE (2005), *heredero* can be an adjective in Spanish.

it represents transfer of wealth (or status) from one generation to the next. Considering endurance, *rico* appeared in four consecutive centuries (1300's-1600's) and its token frequency in the 1600's was above the proposed CIC threshold. The type *por heredero* appeared in the data in only three centuries, never once above the proposed CIC threshold, and *heredero* appeared in the data again only in the 1600's. Because of its endurance and generality, *rico* is proposed to be the central member of its cluster.

Because *heredero* can convey a state that focuses on the subject receiving money and/or a throne, this type is included here. Even if the focus is on a change in which the subject will inherit the throne, the idea of wealth is associated with being king or queen. Example (86) demonstrates how Don Sancho became heir and will become king after his father dies.

(86) *fue concordado que don sancho faria todo seruic'io a su padre en que quedase heredero don sancho el qual nunca se llamo Rey en vida del padre saluo jnfante eredero*
'It was agreed that Don Sancho would be of service to his father upon Don Sancho becoming heir, who was never called King during the life of his father but prince heir.' (*Atalaya de las corónicas.* 15th c. Anon. O'Neill 1999)

The central member of the *pobre* clusters is *pobre* 'poor' and is the only type in these clusters that appears above the CIC threshold. The *pobre* clusters include types, such as *descubierto* 'uncovered' and *desabrigado* 'inadequately dressed', that are symptoms of being poor. In Example (87), the subject not only worries about dying as a result of being left uncovered, but associates this condition with poverty.

(87) *Quando vi tan mal conc'ierto Alto Rey muy soberano Yo me vi del todo muerto Por quedar me descubierto Tanto Lexos del verano Respondi con grand tristeza Como puede ser aquesto Que me dexes Pues que sabes mi pobreza Cesse ya la tu crueza.*

'When I saw such a bad accord, most sovereign High King, I saw myself completely dead from <u>being left uncovered</u> so far from summer. I responded with great sadness. How can it be that you leave me? Since you know of my poverty, stop your cruelty now.' (*Cancionero castellano y catalán de París*. 15th Ctry. Anonymous. O'Neill 1999)

The idea of inadequate clothing as a metaphor for poverty continues in this set of clusters through the *con* + OBJECT generalization. Example (88) depicts a change of state in which the subjects come out with clothing that betrays a social state of poverty.

(88) *E muchas destas van por la calle Arreadas que quando tornan A casa & han tornado A cada qual lo suyo <u>quedan con rropas de asy</u> A tan- da rrotas rraydas & descosydas llenas de suzjedad & mal Apa- reJadas.*
'And many of these go dragging through the street and when they turn toward home, each to her own, <u>they are left with clothes like that</u>: so ripped, torn, unsown, full of filth, and poorly matched.' (*Corbacho*. 15th Ctry. Arcipreste de Talavera. O'Neill 1999)

Example (88), and the other token of *con* + OBJECT, *quedar(se) con poco heredamiento* 'to be left with little inheritance', demonstrate how *con* 'with' can be used to indicate a change of state in which subject comes out with a detraction of physical goods or wealth. This demonstrates the flexibility of the preposition *con* in *quedar(se)* + ADJ in that it can be used in contexts in which its opposite *sin* 'without' could also be used.

The type *quedar(se) sin deuda* in Figure 22 is proposed to intersect with the *rico* 'rich' cluster because it indicates a change in which debt is relieved, thereby allowing the subject to remain with the compensation that would have been used to pay the debt. By being left without debt, the subject is better off although not necessarily rich.

Because of the flexibility of *quedar(se) con* + NOUN and *quedar(se) sin* + NOUN, it is important to consider the meaning of each combination in order to place it in a proposed category structure. The cluster containing the two tokens of *quedar(se) sin* +

OBJECT is placed in a position where it touches both the *pobre* cluster and the *quedar(se) con* + OBJECT cluster in order to indicate semantic similarity. The subject undergoes a change in which they come out with fewer objects than they possessed before, as in Example (89).

(89) *E tantos & tan grandes golpes dio que quebro la acha en piec'as & quedo sin armas.*
'And so many and such great blows gave he that he broke the axe in pieces & he was left without weapons.' (*Oliveros de Castilla*. 1499. Burgos. O'Neill 1999)

5.2.3.4 The *rico / pobre* clusters in the 1500's.

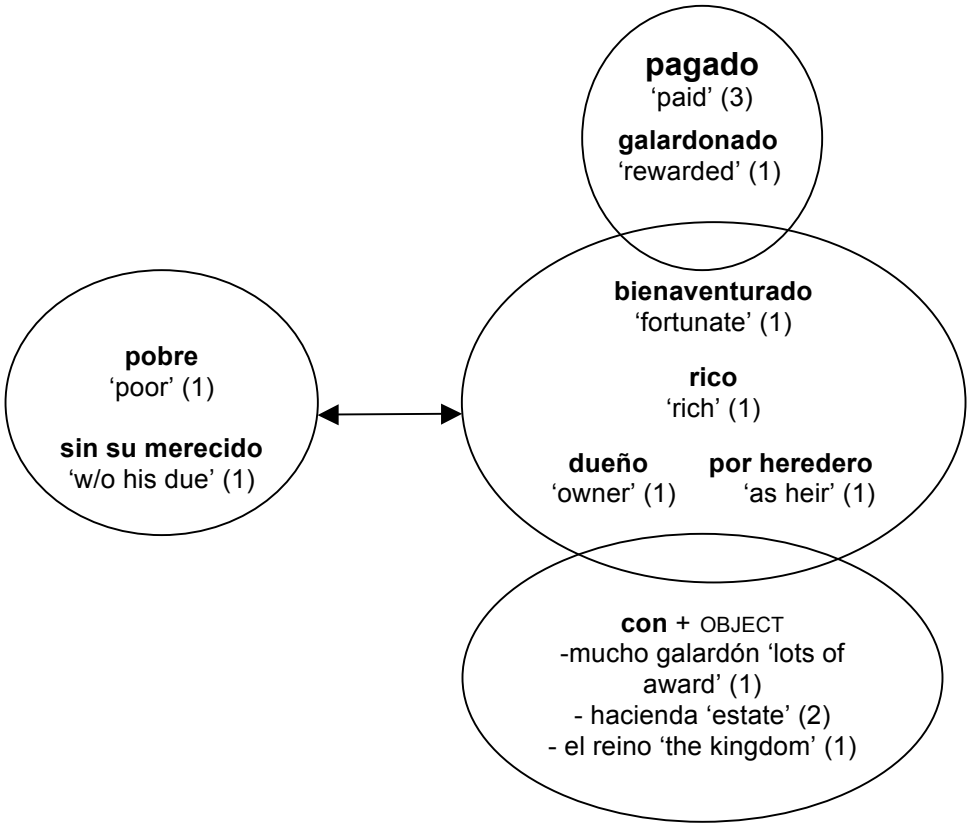


Figure 23, 1500's: The *rico* / *pobre* clusters

As in previous centuries, this set of clusters displays types on both ends of the material wealth continuum; on one side of Figure 23 are the *rico* 'rich' clusters and on the other is the *pobre* 'poor' cluster. Despite the fact that *pagado* 'paid' has more tokens in my data, *rico* is the proposed central member based on the following factors: it has greater endurance, and it has a higher standardized frequency than *quedar(se) pagado* in Davies (2002-) data (*quedar(se) rico*: 0.5 occurrences per million words, *quedar(se) pagado*: 0.1 per million).

In the 1500's, some of the types from the previous *rico* cluster have carried over into this century: *rico* 'rich', *galardonado* 'rewarded', and *por heredero* 'as heir'. The type *dueño* 'owner' presents an interesting case. Even though many of the adjective types included in the data can serve as adjectives or nouns, such as *heredero*, *rico*, or *pobre*, the DRAE (2006) lists *dueño* as a noun only. However, it is included because it is syntactically indistinguishable from an adjective when used in the following token of *quedar(se)* + ADJ shown in Example (90). As Hopper and Thompson (1984: 714) observe, in discourse it is possible to manipulate word class, embedding nouns into adjectival formatives; this is accompanied by symptoms of decategorialization. Evidence of this here is that there is no article and it is used immediately after *quedó descontenta* 'to become unhappy' in which *descontenta* is clearly an adjective. There was a small group of nouns found in the data that appeared in *quedar(se)* + ADJ, embedded as though they were adjectives (See Appendix 1, Section 1).

(90) *Bien sé yo que ella quedó descontenta si no quedó dueña, si, con todo, otro no avía ya madrugado antes.*
'I know well that she would become discontent if she didn't become the owner, if, considering everything, another hadn't already gotten up early before her.' (*Comedia llamada Florinea*. 16th c. (1554). Juan Rodríguez Florián. Lemir)

All of the tokens of *quedar(se) con* + OBJECT in the 1500's refer to a situation in which the subject gains wealth, as in Example (91). This is in contrast to many of the examples in previous centuries in which *quedar(se) con* + OBJECT indicates some type of loss, as in *quedar(se) con poco heredamiento* 'to be left with little inheritance'.

(91) *Por la muerte de su padre Metino quedó con el reino Pigmalión, hermano de la misma Dido, la cual casó con Siqueo.*
'Because of the death of his father, Metino, Pygmalion, brother of the very same Dido, who married Siqueo, was left with the kingdom.' (*Comedia llamada Florinea*. 16th c. (1554). Juan Rodríguez Florián. Lemir)

Looking at the *pobre* 'poor' end of the scale in Figure 23, the only types that have to do with being poor in the 1500's are *pobre* 'poor' and *sin su merecido* 'without his due'. The latter refers to a gardener's pay, as shown in Example (92).

(92) *Y harás poner mejor cobro en la puerta, que el jardinero no quedará sin su merescido.*
'And you will have a better collection place put on the door, so that the gardener won't be left without his due.' (*Comedia llamada Florinea*. 16th c. (1554). J. Rodríguez Florián. Lemir)

The adjective *pobre* 'poor' is not seen again in the data until the 1800's, where it appears only once. In the current century, it is proposed to be the central member because of the fact that it appears in other centuries and because it is more general than *sin su merescido* 'without his due'. In the previous century it appeared above the proposed CIC threshold.

5.2.3.5 The *rico / pobre* clusters in the 1600's.

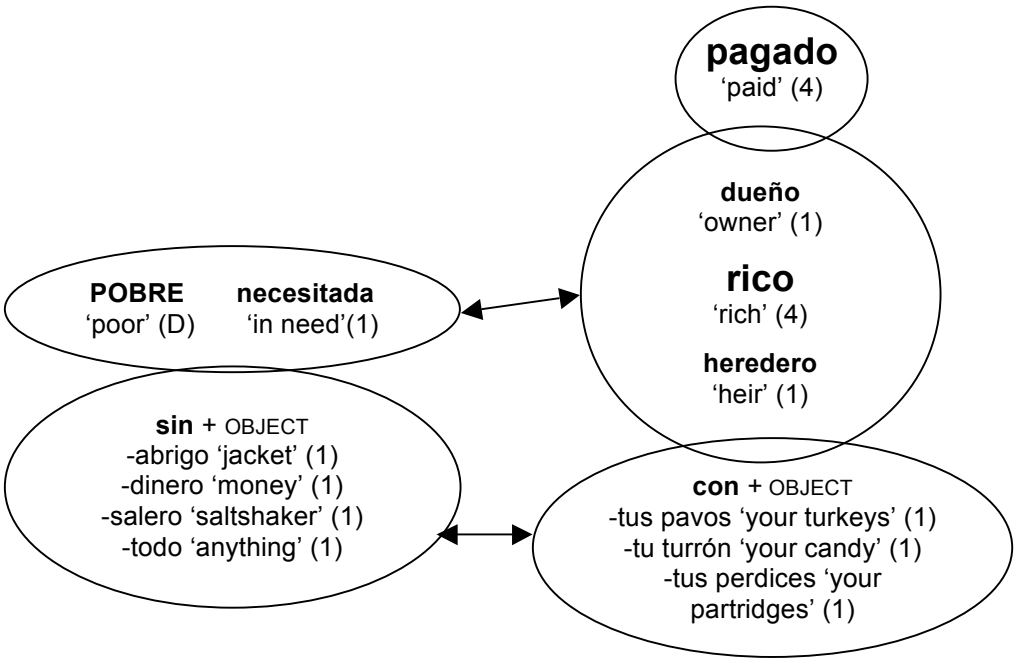


Figure 24, 1600's: Clusters centering on *rico*

Although this set of clusters displays some degree of continuity to the ones from the previous three centuries, a key type was not found in the data: *pobre* 'poor'. As was done in previously (Section 5.2.2.3 with *preso* 'imprisoned', 5.2.2.6 with *obligado* 'obligated' and *libre* 'free', and in Section 5.2.3.1 for *rico* and *pobre*) despite the fact that it didn't appear in my data, *pobre* was included because of its presence in Davies' (2002-) data (see Chart 4, p. 151). Because of its generality and endurance, it is also proposed to be the central member. Nevertheless, the spectrum between wealth and poverty is still present in the adjective types in these clusters. This is also the last century with productive usage of types pertaining to the generalization *quedar(se) con* + NOUN in the data. This century is proposed to mark the beginning of a decline in productivity of the *rico* / *pobre* clusters.

The loss in productivity could partially be attributed to the rise in frequency of other expressions of 'becoming' and the resulting shift in the distribution of adjectives. For example, the adjective *rico* does not appear with *quedarse* in the 20th century data reported on in Bybee & Eddington (2006), but appears two times with *hacerse.* Many of the types from the *rico / pobre* clusters came to appear with *hacer(se)* + ADJ as this form emerged. This is discussed in Section 5.2.3.8 below.

As seen in Figure 24, with four occurrences *rico* 'rich' appears above the CIC threshold in the data providing evidence that it is the central member of the *rico* clusters in this century. However, this is not indicative of an emergent trend in which it goes on to gain token frequency; it is the last time it appears in the data and in the three previous centuries there was only one occurrence per century. The other two types in the *rico* cluster have some continuity to previous centuries. Some form of *heredero* 'heir' appeared in every century since the 1200's and *dueño* 'owner' appeared in the last century and will appear in the prepositional phrase *por dueño* 'as owner' in the next one.

Changes resulting in the loss of specific material goods come from the generalization *quedar(se) sin* + OBJECT. Three of these occurrences come from the same author and are part of a 'multiple' usage of *quedar(se)* + ADJ in which there are seven adjectival complements, or complements of prepositional phrases in this case. A seen in Example (93), there are four instances of *quedar(se) sin* + NOUN marked by numerals in subscript and three instances of *quedar(se) con* + OBJECT marked by letters in subscript. In this 'multiple' there is one occurrence of *quedar(se)* conjugated in the third person singular that extends through the discourse to include a variety of other subjects including 2s, 3p, and 1p.

(93) *Todo lo yerras, señor; mira, si la envías dos pavos, Clara (es más claro que el sol) envía uno a cierta vieja, y otro a cierto Chapetón para coger con el pavo otro regalo mayor: a su agente las perdices; una caja de turrón a una vecina, y la otra a otro solicitador para dar a los que piden de beber la colación; con que tu padre se queda sin salero*$_1$*, tú, señor sin padre*$_2$*, Clara sin todo*$_3$*, y todos, que es lo peor, el uno con tus perdices*$_A$*, la otra con tu turrón*$_B$*, con tus pavos*$_C$ *uno y otro, y sin dinero*$_4$ *tú y yo.*
'You are wrong about everything, sir. Look, if you send two turkeys, Clara (it is clearer than the sun) sends one to a certain old lady, and another to a certain Chapetón in order to get a better gift with the turkey: to your agent the partridges; a box of *turrón* candy to a neighbor, and the other to another solicitor in order to give to those who ask to drink the collation; because of which your father is left without a saltshaker$_1$, you, sir without a father$_2$, Clara without anything$_3$, and all of them, which is the worst, the one with your partridges$_A$, the other with your *turrón* candy$_B$, with your turkeys$_C$ one and the other, and without money$_4$ you and I.' (*Abre el ojo.* 17th c. F. de Rojas Zorrilla. BVMC)

One of the most significant things about this example is that it shows how the author perceives both the structural and semantic similarity of *quedar(se) sin* + NOUN and *quedar(se) con* + NOUN. It provides a colorful example of how these two generalizations are used to indicate a momentous change of state in which many material objects are transferred among many individuals. In this example, the subjects of *quedar(se) con* + OBJECT become better off in terms of material goods and the subjects of *quedar(se) sin* + OBJECT become worse off. In fact the subjects of *quedar(se) sin* + NOUN are so much the worse for the change of state that the father will not even have a salt shaker when it is over. Clara will have nothing, and the fictional speaker and his interlocutor will be left without any money. Even though *pobre* 'poor' does not occur in the data in this century, this example shows how the changes of state described by *quedar(se) sin* + OBJECT can result in poverty.

5.2.3.6 The *rico / pobre* clusters in the 1700's.

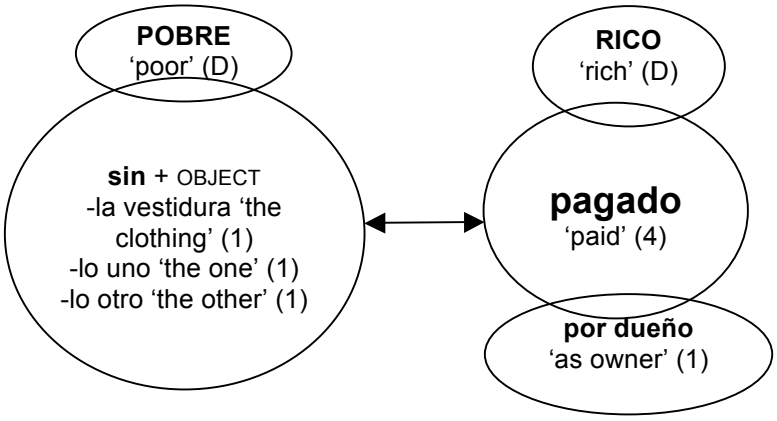


Figure 25, 1700's: The *rico / pobre* clusters.

In the 1700's, as seen in Figure 25, there are notably fewer types in the *rico / pobre* clusters even though the dichotomy of becoming rich or poor is still present. In my own data neither *rico* 'rich' nor *pobre* 'poor' appear in this century and there are no occurrences of *quedar(se) con* + OBJECT that portray changes in wealth. However, *quedar(se) rico* and *quedar(se) pobre* did appear in the *Corpus del español* (Davies 2002-) and are included in this set of clusters to reflect this presence (see Chart 4, p. 174).

The idea of losing clothing as a metaphor of becoming poor continues in this century. In Example (94), the subject is willing to risk losing all of his money and clothes for a chance at winning enough money to buy all new attire. Since losing everything in a dice game would result in the soldier being left naked, the phrase *lo uno y lo otro* is inferred to refer to clothing.

(94) ***Sol****. Yo he hecho mi cuenta, y he menester camisas, jubon, casaca, calçones, médias, çapatos, y sombrero, y con quatro pesos no áy para todo; pues comprar lo uno nuevo, y traer lo otro viejo, no irà bien: quiero jugar, quiçà ganarè para comprarlo todo nuevo.*
Sar. *Y si el dado dize mal?*
Sol. *Me quedare sin lo uno, y sin lo otro, y entonces dire; desnudo nacì, desnudo me hallo, y desnudo morirè.*

'**Soldier**. I have made my count, and I need shirts, soap, jacket, pants, socks, shoes, and a hat, and with four *pesos* there isn't enough for all; well buying the new one, and wearing the old one, won't go together well: I want to play, maybe I'll earn enough to buy everything new.
Sergeant. And if the dice come out badly
Soldier. I shall be left without the one, or the other, and then I will say; naked I was born, naked I find myself, and naked I shall die.' (*Diálogos nuevos en español y en francés*. 18th c. (1708). F. Sobrino. Lemir)

5.2.3.7 The *rico / pobre* clusters in the 1800's.

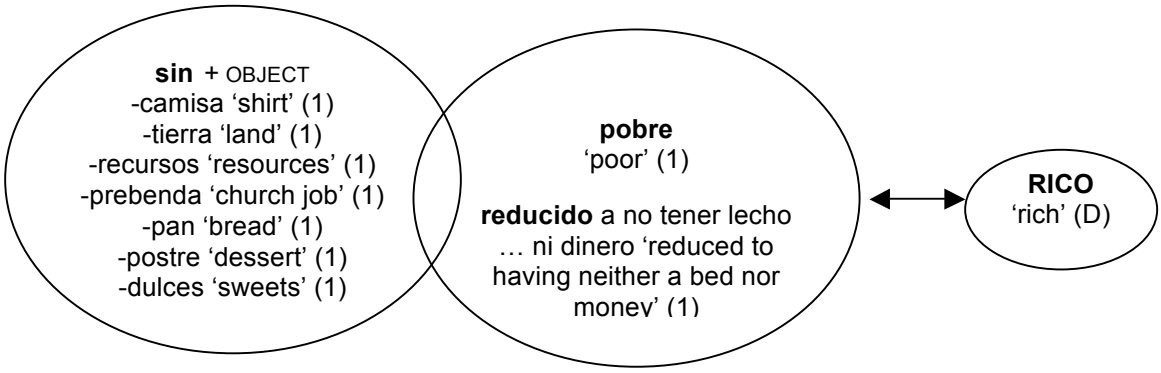


Figure 26, 1800's: Clusters centering on *pobre*.

One of the main things that characterizes the *rico / pobre* clusters in this century is that there were no occurrences of *quedar(se) con* + NOUN in the data. *Rico* 'rich' has not resurfaced in my data, nor are there any types that indicate a change of state in which wealth is gained. Despite this, *quedar(se) rico* does appear in Davies' (2002-) data and is included in these clusters to acknowledge this (see Chart 4, p. 174). The type *pobre* 'poor' does make another appearance in the data and there are other types that indicate a loss of material goods.

The types classified as *quedar(se) sin* + OBJECT shown in Figure 26, above, have to do with a loss of possession. Example (95) demonstrates how a loss of clothing, and of food, are both metaphors for poverty. This is further highlighted by the fact that the

change of state indicated, being left without bread and without a shirt, is one that happens to the 'poor person' (*el pobre)*.

(95) *Por eso la frase sacramental de los desdichados, al verse perseguidos y robados, es ésta siempre: -¿Y qué voy a hacer yo contra ese hombre? ¡Gracias que se ha conformado con esto! Y «esto» es, quizá, el haberse quedado el pobre sin pan y sin camisa.*
'Because of this the sacramental phrase of the unfortunate, upon seeing themselves chased and robbed, is always this: "And what am I going to do against that man? Good thing he has been satisfied with this!" And "this" is, perhaps, the poor person being left without bread and without a shirt.' (*Esbozos y rasguños*. 19th c. (1888). J. M. de Pereda. BVMC)

5.2.3.8 Overview of the *rico / pobre* 'rich / poor' clusters

One of the things that the *rico / pobre* clusters have shown is that there was a group of adjective types in the data that indicated a change of state that left the subject either richer or poorer when used in *quedar(se)* + ADJ. This could be a metaphorical gain or loss, such as *sin pan y sin camisa* 'without bread and without a shirt' in Example (95), or it could be of a specific object such as *sin dinero* 'without money' in Example (93). Even if metaphorical, the change indicates that the subject is either better off or worse off in terms of material wealth.

A salient pattern exhibited in these proposed clusters is that the central members, *rico* 'rich' and *pobre* 'poor', did not have a persistent presence in my data (see Sections 5.2.3.5, 5.2.3.6, and 5.2.3.7); in these cases, they were included because they were present in an outside data source (Davies 2002-). Also, there were cases where the central member did not have the highest token frequency out of all the types in its category. This required other factors to be considered in justifying the proposition that these types were central members such as endurance, generality, and robust presence in Davies' (2002-) data (see Sections 5.2.3.3, and 5.2.3.4). In spite of the need to use a variety of criteria for

determining the central members, these categories expand readily to new types that indicate a loss or gain of wealth. Much of this expansion occurs through the generalized structures *quedar(se) con* + OBJECT and *quedar(se) sin* + OBJECT.

Looking at type frequency, this set of clusters shows an overall pattern in which the category expands and then contracts, much like the *libre* clusters. Table 20, below, shows the total number adjective types in the data and types proposed to belong to the *rico / pobre* clusters; the percentage of all types that belong to the *rico / pobre* is given in parentheses. The types in Table 20 are either in the *rico* and *pobre* clusters, or are types that intersect with these (e.g. types from *quedar(se) con* + OBJECT or *quedar(se) sin* + OBJECT when these describe a gain or loss of possession). Types inserted into these clusters from Davies' (2002-) data were not considered. The lack of figures from the 1900's is shown to emphasize the fact that Bybee & Eddington (2006) did not find any of these types in their data occurring with *quedar(se)*.

Table 20. Types in the *rico / pobre* clusters.

	Total # types	Types in *rico / pobre* clusters
1200's	29	1 (3.5%)
1300's	40	7 (17.5%)
1400's	121	14 (11.5%)
1500's	153	11 (7%)
1600's	171	12 (7%)
1700's	135	5 (3.5%)
1800's	141	9 (6.5%)
1900's*	54	--

* Figures from Bybee & Eddington (2006)

In the early centuries, types in the *rico / pobre* clusters go from accounting for 3.5% of all types in the 1200's to accounting for 17.5% of all types in the 1300's. By the 1400's types in the *rico / pobre* clusters account for 11% of all types. The type frequency

diminishes from the 1400's to the 1800's in which they end up accounting for 6.5% of all types. Furthermore, the clusters from the 1700's and the 1800's show a loss of some key types. The type *rico* 'rich' doesn't appear in my data after the 1600's. In the 1700's and 1800's there were no occurrences of *quedar(se) con* + NOUN that reflected a change in the subject's status of wealth. Also, in Bybee & Eddington's (2006) data from the 1900's, there were no occurrences of *quedar(se) con* + NOUN documented whatsoever.

In the data, the cumulative token frequency of the types in the *rico / pobre* clusters increases over the centuries, and then decreases. Looking at each century, Table 21 shows the total number of tokens in these clusters, the tokens taken from the types represented in Table 20, and gives the percentage of all tokens that these represent. The third column provides the standardized overall frequency of *quedar(se)* + ADJ in my data as occurring per 10k words.

Table 21. Tokens in the *rico / pobre* clusters.

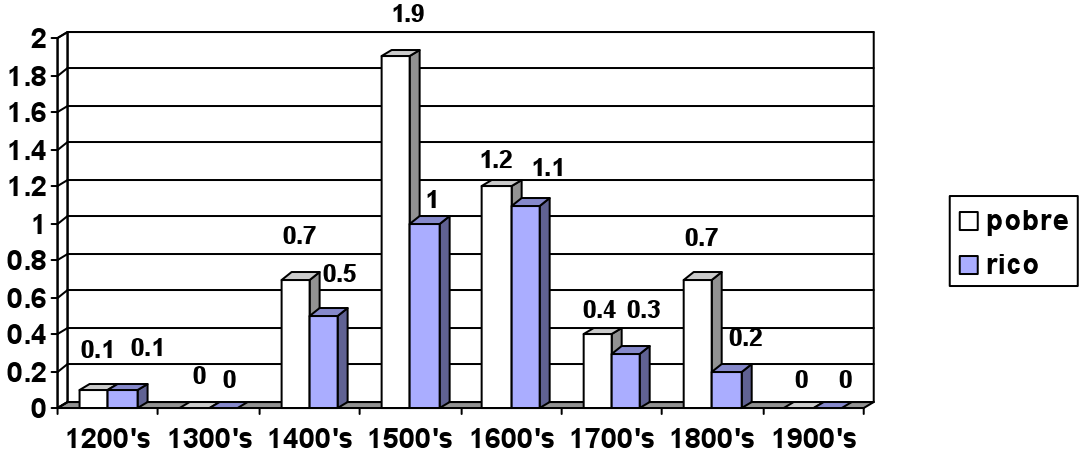
	Total # of tokens	Tokens in *rico / pobre* clusters	# *qdr(se)* + ADJ per 10k words
1200's	43	3 (7%)	0.42
1300's	76	7 (9%)	0.96
1400's	196	17 (8.5)	1.18
1500's	271	11 (4%)	4.85
1600's	280	13 (4.5%)	6.00
1700's	253	4 (1.5%)	5.82
1800's	255	9 (3.5%)	2.64
1900's*	181	--	1.83

* Figures from Bybee & Eddington (2006)

The diachronic pattern in the token frequency in the *rico / pobre* clusters resembles the one related to type frequency (Table 20). In the 1300's, tokens from the *rico / pobre* clusters reach a peak, accounting for 9% of all tokens found in the data for that century. There is a noticeable decline in the portion of tokens accounted for by types

in the *rico / pobre* clusters after the 1400's; after this century, these tokens will never account for more than 4.5% of all tokens in the data. Looking at the third column, as the overall frequency of the change-of-state construction *quedar(se)* + ADJ rises in the 1500's and 1600's, the token frequency of the *rico / pobre* clusters generally diminishes as measured by the percentage of all tokens that they account for.

Chart 4. Overall frequency of *quedar(se) pobre* & *quedar(se) rico* in 8 centuries of the *Corpus del español* (Davies 2002-); occurrences per 1,000,000 words.



A similar pattern is seen in the *Corpus del español* (Davies 2002-). Looking only at the types *quedar(se) pobre* 'to become poor' and *quedar(se) rico* 'to become rich' in the *Corpus del español* (Davies 2002-), the token frequency rises and then diminishes. Chart 4 shows the number of occurrences per one million words of *quedar(se) pobre* and *quedar(se) rico* from the *Corpus del español* (Davies 2002-). Figures from the 1900's come from the fictional register. The search rendered all conjugations of the verb

quedar(se) as it appeared next to all forms (masculine / feminine, singular / plural) of the adjectives *pobre* and *rico*[26].

Both *quedar(se) pobre* 'to become poor' and *quedar(se) rico* 'to become rich' begin with low token frequency in the 1200's (0.1 occurrences per 1 million words each) and then gain in token frequency before tapering off. *Pobre* reaches its peak in the 1500's at 1.9 occurrences per 1 million words and *rico* reaches its peak in the 1600's at 1.1 occurrences per 1 million words. Compared to other central members, these peaks in standardized overall frequency were not as high; in Davies' (2002-) data *quedar(se) libre* reached a peak of 10 occurrences per million words in the 1600's, and *quedar(se) solo* reached a peak of 18.6 occurrences per million words in the 1900's fictional register. After the peak, there was a general reduction in token frequency. In the 1900's, there were no tokens of *quedar(se) rico* or *quedar(se) pobre* in the fictional register of the *Corpus del español* (Davies 2002-) and the tokens in Chart 4.

Taken as a whole, the figures from Table 21, Table 20, and Chart 4 provide evidence that, as the category expanded in type frequency, there was correlating increase in token frequency. As the category lost types, the token frequency of the category also diminished. Furthermore, the central members *quedar(se) pobre* and *quedar(se) rico* underwent a rise and fall in token frequency as well.

Another factor that could have played a role in the category reduction observed in the *rico / pobre* clusters was the emergence of other verb + adjective expressions of 'becoming'. In the 1900's, Bybee & Eddington (2006) found two occurrences of *hacerse rico* 'to get rich' but no occurrences of *rico* with any other verb. A search on the *Corpus*

[26] Individual tokens were analyzed to confirm that they expressed a change of state with an animate subject. Two tokens of *quedar(se)* with *rico* expressed remaining: one in the 1500's, and one in the 1200's (the only one). All occurrences of *quedar(se)* with *pobre* expressed a change of state.

del Español (Davies 2002-) reveals that while combinations of *quedar* and *rico* decline from the 1600's to the 1900's (from 1.1 to 0.2 occurrences per million words), combinations of *hacer* and *rico* rise steadily from 2.6 to 4.2 occurrences per million words in this same time span. However, one of the differences between *quedar(se) rico* and *hacer(se) rico* is that the latter can be used transitively with an object pronoun (*te voy a hacer muy rico* 'I'm going to make you very rich') and intransitively with a 'reflexive' pronoun (*te vas a hacer muy rico* 'you will get rich'). Both depict a change of state but the intransitive usage is the one considered by Bybee & Eddington to be a verb of 'becoming' similar to *quedar(se)* + ADJ. These were more frequent; in the 1800's, only seven out of 69 occurrences in Davies' (2002-) were transitive. A search for *hacer(se) pobre* 'to become poor' revealed that these declined in Davies data going from 2.2 occurrences per million words in the 1600's to 0.1 per million in the 1900's fictional genre. Even though it was losing frequency, it was more frequent than *quedar(se) pobre* which went from 1.9 occurrences per million words in the 1600's to zero in the 1900's. Further investigation is merited in order to see how other adjectives interacted with *quedar(se)* and *hacer(se)*. At the very least, this search on the *Corpus del Español* (Davies 2002-) shows that as far back as the 1200's, *rico* and *pobre* were being used concurrently with both *hacer(se)* and *quedar(se)* to express a change of state.

Table 22 shows that there were several types that appeared in the data in at least three centuries. The rows corresponding to *sin* + OBJECT and *con* + OBJECT reflect the types that intersected with the *rico* or *pobre* clusters and, therefore, appear in contexts that portray a loss or gain of material wealth. Despite the fact that the central members *rico* and *pobre* showed relatively low token frequency compared to other central

members and inconsistent usage in my data, they each appear from the 1200's to the 1800's when Davies' (2002-) data is considered.

Table 22. Types from the *rico / pobre* clusters that appear in at least 3 centuries.

	1200	1300	1400	1500	1600	1700	1800	1900
rico 'rich'	DDD	====	====	====	####	DDD	DDD	
pobre 'poor'	DDD	====	####	====	DDD	DDD	====	
pagado 'paid'		====	====	====	####	####		
por heredero 'as heir'	====		====	====				
sin + object			====	====	====	====	====	
con + object		====	====	====	====			

(====) = presence of type in data below CIC threshold
(####) = presence of type in data above CIC threshold
(DDD) = presence only in Davies *Corpus del español* (2002-)

One of the reasons that *quedar(se) rico* 'to become rich' and *quedar(se) pobre* 'to become poor' don't appear in my data in every century is due to the fact that they are not as robust as other central members. As mentioned above, on a century-by-century basis they have a lower standardized overall frequency than *quedar(se) solo* 'to be left alone', and *quedar(se) libre* 'to be set free' in the *Corpus del español* (Davies 2002-) and also have a lower relative frequency than these types in my data. In spite of this difference in frequency, the category shows evidence of an emergent gain in token and type frequency followed by a contraction of these. As the token frequency of *quedar(se) rico* and *quedar(se) pobre* wanes, related adjective types are discontinued in usage until the entire category seems to have disappeared in written data by the 1900's. However, even though the central members were not as robust as ones from other categories, the *rico / pobre* clusters represent a plausible category of adjectives that were used in *quedar(se)* + ADJ for seven centuries (i.e. the 1200's to the 1800's). Even low frequency formulaic sequences show endurance as do the categories to which they belong and it is possible to apply the same kind of analysis here as applied to previous clusters. These clusters

provide an ideal set of data with which to study the relationships of opposites. The MDS analysis (Ch. 6) shows that opposites are perceived as more similar than adjectives that are unrelated semantically.

5.3 Changes in mental state

This section examines clusters of adjectives that, used in *quedar(se)* + ADJ, mainly describe changes in mental states. The clusters analyzed here are referred to as the *alegre / satisfecho* 'happy / satisfied' clusters and are presented in Section 5.3.1. One of the reasons to include the analysis of this set of clusters was to reveal diachronic tendencies that haven't been fully explored yet. First, this set of clusters has many marginal types that may be related only through a chain of family resemblance, as was depicted in Figure 5, p. 101. A similar chain of family resemblance linking disparate types such as *muerto* 'dead' and *contento* 'happy' was proposed in Bybee & Eddington (2006: 334). One of the motivations of the present analysis is to better understand the series of links that they proposed. Second, this set of clusters shows how a central member may change over time, between *alegre* and *satisfecho*. In addition to these clusters, three other sets of clusters of adjectives that describe mental states are included in Appendix 1: the *triste* 'sad' clusters (Section 2), the *confuso / sorprendido* 'confused / surprised' clusters (Section 3), and the *convencido* 'convinced' clusters (Section 4). There is an antonymic relationship between the *alegre / satisfecho* clusters and the *triste* clusters that could render an analysis similar to the one of the *rico / pobre* clusters in the previous section. However, in order to make a theoretical point, the analysis of this set of clusters focuses on the relationships of family resemblance and on the changing central members.

5.3.1 Happiness and its extensions: the *alegre / satisfecho* 'happy / satisfied' clusters

In each century, this set of clusters has been organized around a central category of adjective types whose central member describes a state of happiness. Since the central member of this set of clusters changes over time, they will be referred to as the *alegre / satisfecho* 'happy / satisfied' clusters in order to acknowledge this mutation and for the sake of convenience. This set of clusters will gain in type frequency as time goes on and some of the marginal adjective clusters are only related to the central one through a series of linked clusters in a chain of family resemblance.

5.3.1.1 The *alegre / satisfecho* 'happy / satisfied' clusters in the 1200's

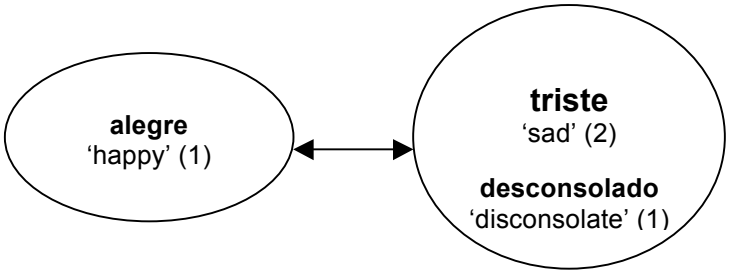


Figure 27, 1200's: The *alegre / satisfecho* clusters.

As seen in Figure 27, there are two clusters that are related because they contain types that are opposites: *alegre* 'happy' in one, and *triste* 'sad' and *desconsolado* 'disconsolate' in the other. Unlike previous sets of clusters in which there were groups of opposite clusters presented in the same figure, the figures depicting the *alegre / satisfecho* clusters will only show opposites in the 1200's (Figure 27) in order to show that this dichotomy exists. However, no opposites will be presented in the *alegre / satisfecho* clusters beyond the 1200's because this type of relation has already been explored.

Figure 27 shows all types in the 1200's that, used in *quedar(se)* + ADJ, indicate a change of state in which the subject becomes either happy or sad. Despite the fact that *triste* 'sad' occurs twice, none of these types occur in the data above the CIC threshold. It is a testament to its endurance that *alegre* appears in this century, the earliest century from which data was collected. Even though the central member of this set of clusters may change over time, and the category may expand, the original central member, *alegre,* appears very early in the data and is proposed to serve as a model for extension. In Example (96), we see that the subject became happy as a reaction to having stated 'this' and turning toward the *villa*.

(96) *Quando esto ouo dicho tornose para la villa: & Boymonte quedo muy alegre.*
'When he had said this he turned toward the *villa*: and Boymonte became very happy.' (*Gran Conquista de Ultramar*.' 13th c. Anon. O'Neill 1999)

Similarly, Example (97) shows how the subjects became sad as a reaction to the departure of a member of their social group. This example also has some of the properties of numerous early example of *quedar(se) solo* 'to be left alone' in which both 'remaining' and 'becoming' are present. The one person leaves and the others remain behind, being left sad. Also, this example shows how the change-of-state expression *fincar(se)* + ADJ (*fincauan muy alegres* 'they became very happy') could be used concurrently with *quedar(se)* + ADJ in a very small span of discourse.

(97) *le dio su bendicion ala partida:& ellos quedaron muy tristes porque se yua: ca no sabian quando se tornaria E fincauan muy alegres porque tenian esperana en dios.*
'He gave him his blessing at the departure and they became very sad, because he was going, because they did not know when he would return. And they became very happy because they had hope in God.' (*Gran Conquista de Ultramar*. 13th c. Anon. O'Neill 1999)

5.3.1.2 The *alegre / satisfecho* 'happy / satisfied' clusters in the 1300's

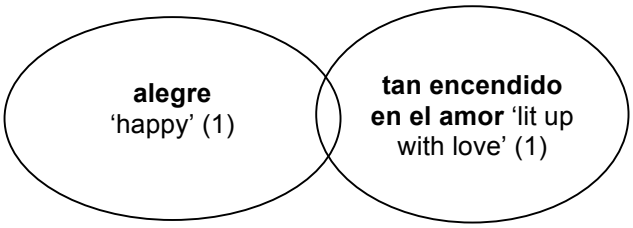


Figure 28, 1300's: The *alegre / satisfecho* clusters.

Figure 28 shows that, although no types occur above the CIC threshold, the type *alegre* 'happy' continues to appear in the data. This is the first century in which types having to do with love appear in the data, as in Example (98). These 'love' types will continue through the 1700's and are proposed to relate to *alegre* because it usually denotes a positive change of emotional state.

(98) *En esto desperto vlixes & quedo tan encendido enel amor de quella vision.*
'In this Ulisses woke up and was so lit up with love from that vision.'
(*Sumas de la historia troyana*. 14th c. Leomarte. BVMC.)

5.3.1.3 The *alegre / satisfecho* 'happy / satisfied' clusters in the 1400's

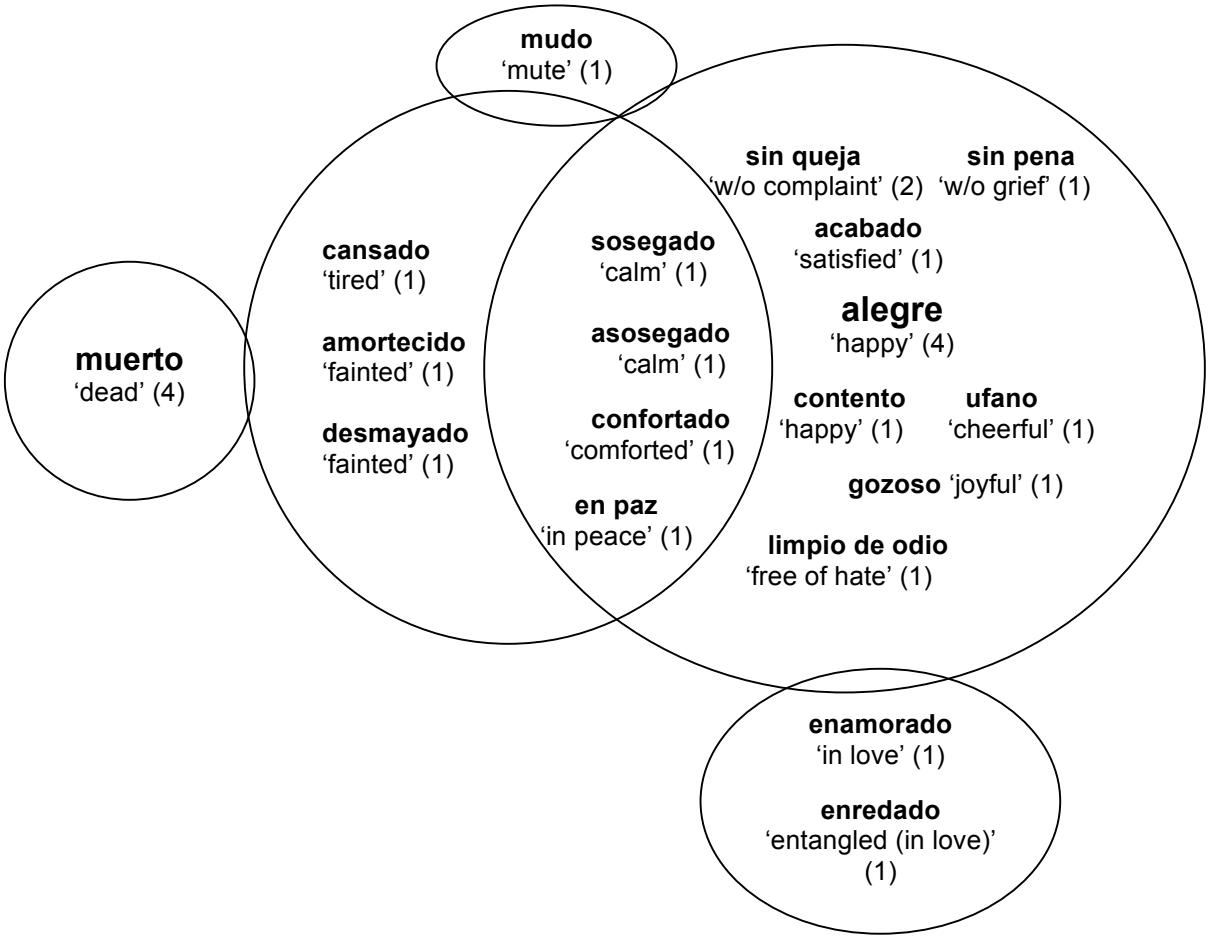


Figure 29, 1400's: The *alegre / satisfecho* clusters.

One of the aims of the analysis of the proposed *alegre / satisfecho* clusters in this century is to show how categories can extend to the point where seemingly unrelated adjectives, such as *muerto* 'dead' and *alegre* 'happy', are linked through intermediary adjective types. Regarding the proposed *alegre / satisfecho* clusters from this century forward, there are some issues to take into account. First, *muerto* 'dead' is the central member of a set of increasingly productive clusters that are shown in Appendix 1

(Section 5). However, it is one of the links in the chain of family resemblance proposed in Bybee & Eddington (2006) and is shown here for that reason. Since the *muerto* clusters are shown with more detail in the appendix, only *muerto* is shown here. The second issue is that *mudo* 'mute' is proposed to be most closely related to *sosegado* 'calm', *asosegado* 'calm', *en paz* 'in peace', and *confortado* 'comforted'. Both of these proposed pathways of linkage were later supported by the MDS study in Chapter 6:.

The central member of this set of clusters in the 1400's is *alegre* 'happy' which, other than *muerto*, is the only type that appears above the CIC threshold. Types that are proposed to be related to *alegre* are placed in the same cluster, represented by the circle in which it occupies a central position. Types that have to do with being 'happy' are placed toward the bottom, i.e., *contento* 'happy' (Ex. (99)), *gozoso* 'joyful' (Ex. (100)), and *ufano* 'cheerful'. Types that are more directly synonymous with being 'satisfied' are placed toward the top of the cluster: *acabado* 'satisfied' (Ex. (99)), *sin pena* 'without grief', and *sin queja* 'without complaint' (Ex. (101), no. 3). Evidence of relatedness would later be confirmed by the similarity study in the present investigation (Chapter 6). Even though *alegre* 'happy' was not one of the types included in the similarity study, synonymous types (e.g. *contento* 'happy' and *encantado* 'enchanted') were shown to be perceived as very similar to *satisfecho* 'satisfied'. Other justification for placing these types in the same circle, thereby highlighting the proposal that they are semantically similar, comes from the adjective placement done by Bybee & Eddington's (2006: 334) native speaker consultant. She placed *satisfecho* 'satisfied', *a gusto* 'pleased', and *contento* 'happy' in close enough proximity that the investigators enclosed these types in a bubble.

(99) *dixo el moc'uelo con cara alegre yo soy E luego desaparec'io E ella quedo muy acabada e contenta de gozo e maraujllosa deuoc'ion.*
'With a happy face the young boy said "I am." And then he disappeared. And she was left very satisfied and happy with joy and marvelous devotion.' (*Exemplario por ABC*. 15th c. Anon. BVMC.)

(100) *pusolo en vn ar- ca & cerrola conla llaue & quedo ella muy gozosa & alegre dizien- do que ya tenia buenos rrehenes.*
'[She] put it in a chest, and locked it with a key, and she became very joyful and happy saying that now she had good hostages.' (*Meditations of Pseudo-Augustine (olim. De infantia Salvatoris)*. 15th c. Anon. BVMC.)

(101) *Señora, este es otro e segundo punto, el qual si tú con tu mal sofrimiento no consientes, poco aprouechará mi venida e, si, como prometiste, lo sufres, tú quedarás sana*$_1$ *e sin debda*$_2$ *e Calisto sin quexa*$_3$ *e pagado*$_4$.
'Ma'am, this is another and second point, that if you with your terrible suffering do not consent, you will not make the most of my coming and, if, as you promised, you suffer through it, you will come out healthy$_1$ and without debt$_2$ and Calisto without a complaint$_3$ and paid$_4$.' (*La Celestina*. 15th c. F. de Rojas. BVMC.)

The three types that denote a sense of calmness, *sosegado* 'calm', *asosegado* 'calm', and *confortado* 'comforted' (Ex. (102)), are proposed to intersect with the *alegre* cluster even thought they may be perceived as also being related to *cansado* 'tired', *amortecido* 'fainted', and *desmayado* 'fainted' (Ex. (103)). These 'calm' types are proposed to intersect with the *alegre* cluster because, used in *quedar(se)* + ADJ, they describe a change of state in which the subject finds themselves in a positive emotional state. This becomes clearer in upcoming centuries.

(102) *E despues que el apostol todo esto ouo dicho fuese de delante del E el Rey don Ramiro quedo muy confortado.*
'And after the apostle had said all of this he went before him. And the king Don Ramiro became very comforted.' (*Suma de las corónicas.* 15th c. Anon. BVMC)

(103) *Cubriose del escudo que non se paresc'ia nada e fuese para ela & ela ujolo venjr & commo ela se miro enel escudo & se vio tan disforme quedose amortec'ida & vino jason & degolola con su espada.*
'He covered himself with his shield so that nothing was showing and he went toward her. And she saw him come and as she saw herself in the

shield, and she looked so deformed, she fainted. And Jason came and slit her throat with his sword.' (*Exemplario por ABC*. 15th c. Anon. BVMC)

Israel (1996), in his study of the *way* construction, argued that novel occurrences appeared through analogical extension to existing, established types. This notion has been supported by other researchers using the exemplar model (Bybee 2006, Bybee & Eddington) and by researchers of construction grammar (Goldberg 1995, 2006). Not only is *alegre* 'happy' a type that has occurred in the data since the 1200's, but it occurs with enough token frequency in the 1400's for it to be considered a CIC in this study. Looking at the data up to the present century, we see that *alegre* 'happy' was the only type that occurred in this set of clusters in the 1200's. In the 1300's it was the only type other than *encendido del amor* 'lit up with love'. Looking at Figure 29, we see that there is an expansion of each of these types from one century to the next, giving us the types found in the data in the 1400's. There are some types more closely related to *alegre* and others more closely related to the emergent type *satisfecho* 'satisfied', as mentioned previously in this section (e.g. *acabado* 'satisfied', *sin pena* 'without grief', and *sin queja* 'without complaint'). Furthermore, there are two 'love' types: *enamorado* 'in love', and *enredado* 'entangled (in love)'.

New types emerge in the data that could have arisen through extension to *sosegado* 'calm', *asosegado* 'calm', and *confortado* 'comforted' because they show a lack of physical movement (e.g. *cansado* 'tired', *amortecido* 'fainted', and *desmayado* 'fainted'). However, they could have also arisen through analogical extension to *muerto* 'dead' which was found in the data in the 1200's but not in the 1300's. In fact, looking at *quedar(se) amortecido* 'to faint' in Ex. (103), it is more of a physical reaction in which the Medusa sees herself and falls to the ground fainted. What is important here is that

there are two plausible pathways of analogical extension that could account for the emergence of types such as *cansado* 'tired', *amortecido* 'fainted', and *desmayado* 'fainted'. Furthermore, it is also plausible that the two pathways worked together.

It seems that it is often the case that extension is based on similarity to high-frequency types, or, at least, to types that appear previously in the data. As Bybee & Eddington (2006: 329) state, their goal was to "produce a conceptual clustering that can predict subsequent uses of the constructions". If the data from the *alegre / satisfecho* clusters from the two previous centuries were used to predict subsequent usage, it is conceivable that the types *sosegado* 'calm', *asosegado* 'calm', and *confortado* 'comforted' came to be used in *quedar(se)* + ADJ based on analogy to *alegre* or other types in these clusters.

In that vein, it is also conceivable that the types *cansado* 'tired', *amortecido* 'fainted', and *desmayado* 'fainted' could have emerged through analogy to *quedar(se) muerto* 'to be left dead / to be killed'. *Muerto* 'dead' first appeared in the data in the 1200's with one occurrence and was proposed to be surrounded by types relating to being hurt (see Appendix 1, Section 5). Although it did not appear in these data in the 1300's, it did appear with the verb *quedar(se)* once in the *Corpus* (Davies 2002-). There was an ensuing rise in the token frequency of *quedar(se) muerto* as evinced by the fact that it appeared as a CIC in the 1400's. Also, the *muerto* clusters expanded in the 1400's. Since the type *dormido* 'asleep', which goes on to be a robust CIC, most likely a prefab, hadn't appeared in these data or Davies' (2002-) yet, it is reasonable that *muerto* could have served as a model of analogy somewhere along the chain of extension that resulted in the production of *cansado* 'tired', *amortecido* 'fainted', and *desmayado* 'fainted' in the

expression *quedar(se)* + ADJ. After all, Example (103) above, demonstrates how *quedar(se) amortecido* 'to faint' resulted in the subject being killed. There is one more point regarding the type *muerto*; it was not included in the *alegre / satisfecho* clusters in the 1200's because there were no plausible intermediary types in the data that could have linked these.

5.3.1.4 The *alegre / satisfecho* 'happy / satisfied' clusters in the 1500's

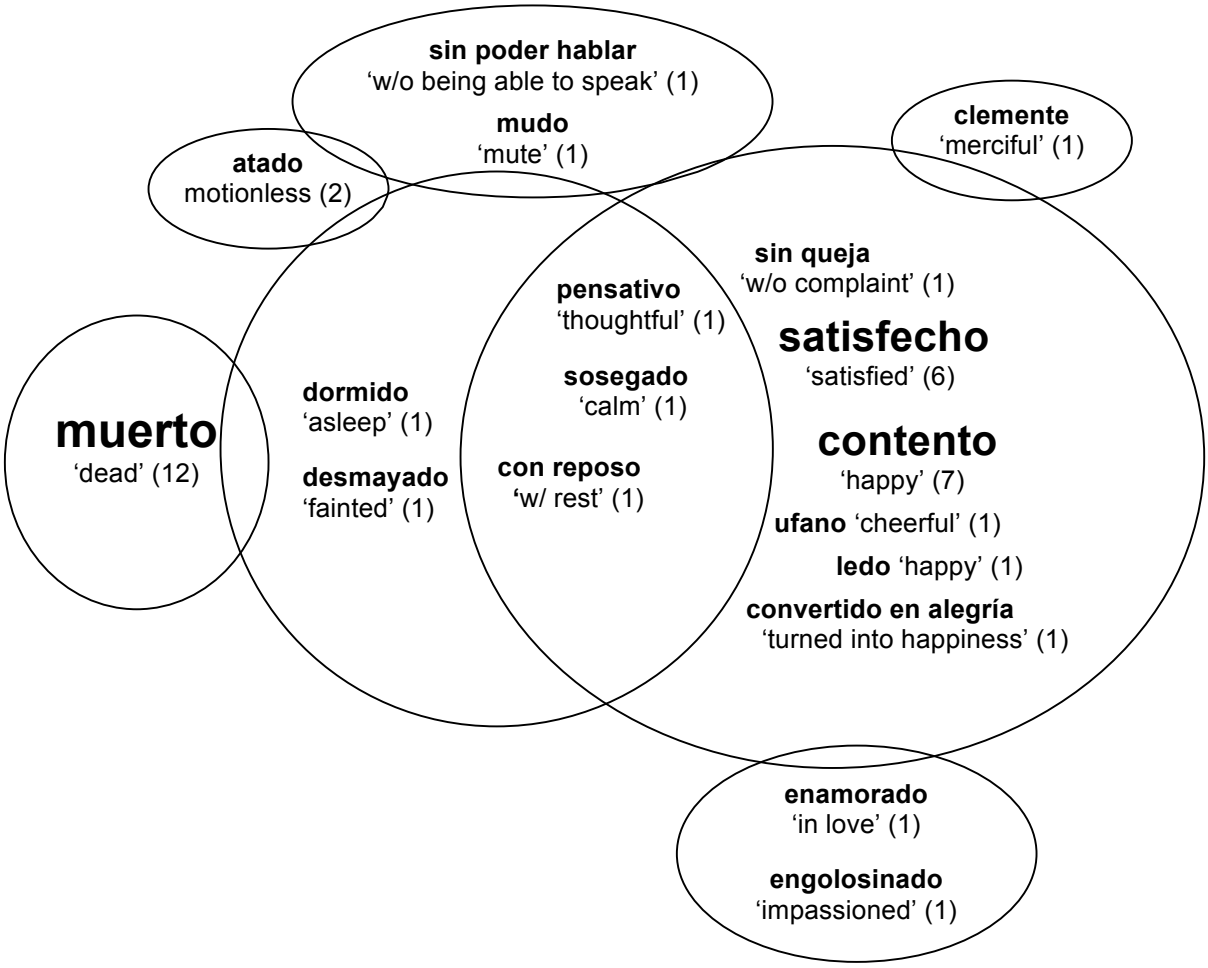


Figure 30, 1500's: The *alegre / satisfecho* clusters.

In the 1500's, the *alegre / satisfecho* clusters show an expansion of types relating to *alegre* 'happy' and *satisfecho* 'satisfied' while maintaining a similar proposed

structure to the previous century. Types that will later gain significantly in token frequency emerge in this century. This is the first century in which the type *quedar(se) satisfecho* 'to become satisfied' (Ex. (104))appears in the data, doing so with enough tokens (6) to put it above the CIC threshold of four. Although *dormido* 'asleep' also appears for the first time in the data, there was only one occurrence (Ex. (105)) and this is also the first century in which it is found in Davies' (2002-) data. The type *atado* (Ex. (106)) is used in this century to convey motionless.

(104) *Pues eso conocéis, dad ahora gracias a Dios por ello, y procurad de no serle ingrato, y pues vos quedáis satisfecho.*
'Well you were familiar with that, give thanks to God now for it, and try not to be ungrateful to him, and then you will be satisfied.' (*Diálogo de las cosas acaecidas en Roma.*16th c. A. de Valdés. BVMC)

(105) *...luego que cayó en la cama quedó dormido.*
'... after he fell in the bed and fell asleep.' (*Sobremesa y alivio de caminantes.*16th c. (1569). J. Timoneda. BVMC)

(106) *que muchos a avido que perdieron sus naturales fuerças de sólo ver a sus amigas, quedando como atados. E muchos, no pudiendo moverse, estuvieron parados y quedos e algunos, tropeçando y temblándoles las piernas, cayeron en tierra, otros han perdido el habla.*
'There have been many who lost their natural strength just from looking at her friends, being left as thought tied. And many, not being able to move, were on their feet and still. And others, tripping with their legs trembling, fell to the ground. Others have lost their speech.' (*Laberinto de amor.*16th c. (1546). J. Boccaccio. BVMC)

(107) *y avemos de mirar que élla queda contenta de tu conversación y gentileza.*
'And we must recognize that she becomes happy with your conversation and kindness.' (*La Comedia Serafina.* 16th c. (1521). Anon. Lemir)

Despite the emergence of important new types, the central member from the previous century, *alegre* 'happy', did not appear in this century in my data. Instead, *contento* 'happy' (Ex. 107), which had only one occurrence in the previous century, and *satisfecho* 'satisfied', which had none, have both gained in token frequency and are

candidates for being the central member of the category. In Davies' (2002-) data, *quedar(se) satisfecho* had a higher standardize frequency (5.5 occurrences per million words) than *quedar(se) contento* (3.5 per million), and will have a higher standardized frequency in subsequent centuries. This, combined with the fact that it is clearly more frequent in upcoming centuries than *contento* leads to the proposal that *satisfecho* is the central member.

5.3.1.5 The *alegre / satisfecho* 'happy / satisfied' clusters in the 1600's

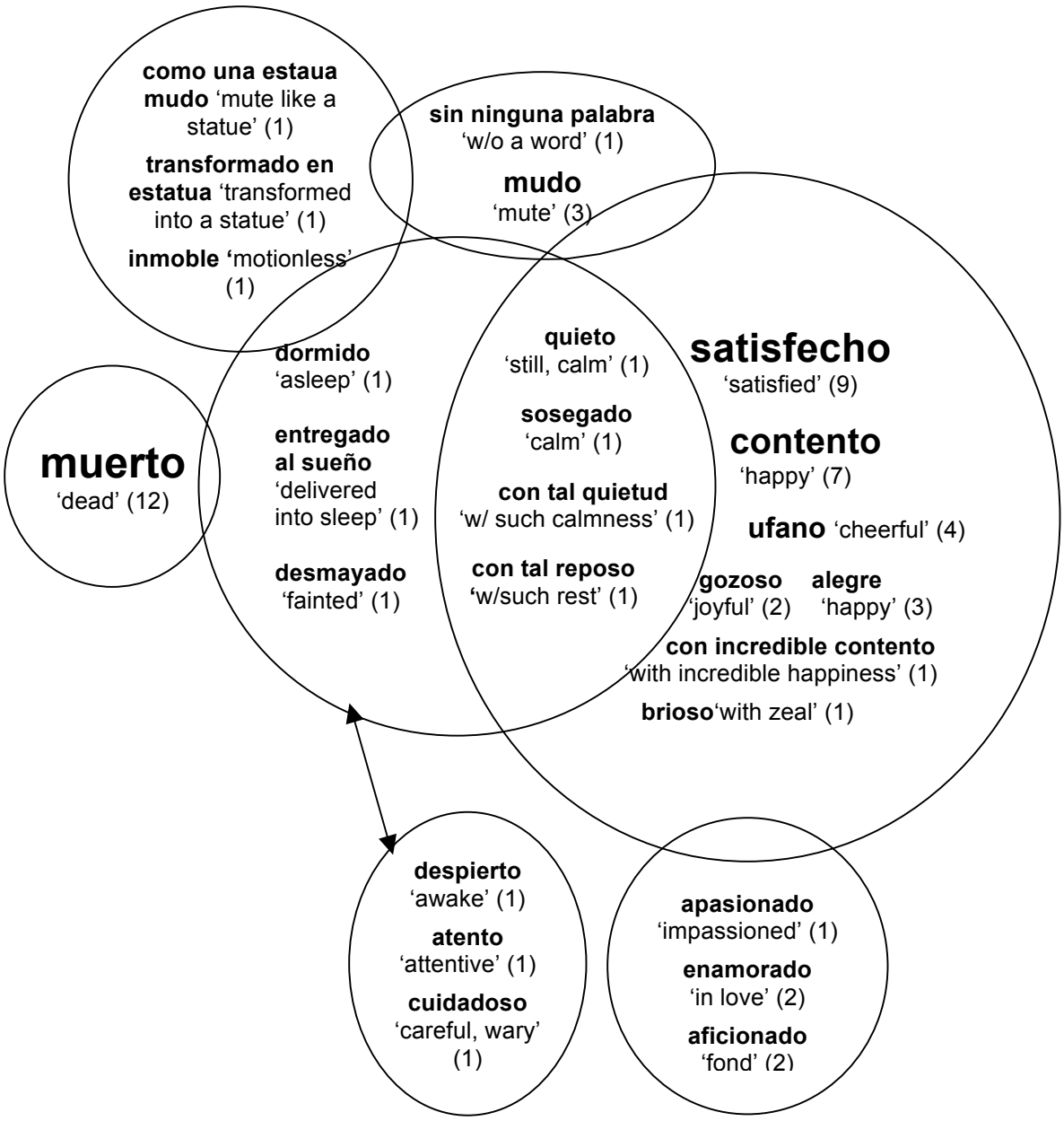


Figure 31, 1600's: The *alegre / satisfecho* clusters.

The proposed category structure depicted in Figure 31 maintains the overall shape from the two previous centuries. With nine occurrences in the data, *quedar(se) satisfecho* 'to become satisfied' has surpassed *quedar(se) contento* 'to become happy' in relative token frequency. Based on token frequency, *satisfecho* 'satisfied' is the proposed central

member of the *alegre / satisfecho* clusters in this century and in the three that follow. This slight difference in frequency could be indicative of the finding in Davies' (2002-) data that the difference between *quedar(se) satisfecho* and *quedar(se) contento* is even greater than in the century before (*quedar(se) satisfecho*: 6.7 occurrences per million words, *quedar(se) contento*: 2.1 per million).

Two new types appear in the data that will go on to gain considerably in token frequency: *inmoble* 'motionless' (Ex. (108)), and *quieto* 'still, calm'. *Inmoble* is an older form of *inmóvil* 'motionless' that means exactly the same thing and even has a similar phonetic shape. It will be one of the types with the highest token frequency in the 1900's (Bybee & Eddington 2006). Taken out of context, *inmoble* (and later *inmóvil*) would seem to be most closely related to types such as *quieto* that portray a lack of motion. Certainly this is a viable connection and was deemed to be the most plausible one by Bybee & Eddington (2006: 333); "Since *quieto* indicates stillness, quiet, and inferentially peacefulness, each of the other adjectives share features with it: *callado* (absence of sound), *inmóvil* (absence of motion), and *tranquilo* (peacefulness)". At the same time, when examining the context, *inmoble* portrays a reaction of surprise to an unexpected situation, as in Ex. (108). In these data *inmóvil* was found to almost always describe a change of state brought about in the subject through being scared or surprised. Because of this it is proposed to belong to the *suspenso / confuso* clusters as well (see Appendix 1, Section 3).

(108) *Inmoble se quedó Mitilene de lo que le refirió su tío, y ... no supo con el susto responderle.*
'Mitilene became motionless from what her uncle told her, and... with the fright she didn't know how to respond.' (*Los dos soles de Toledo.* 17th c. (1521). A. Alcalá y Herrera. BVMC)

The types *despierto* ‘awake’ (Ex. (109)), *atento* ‘attentive’, and *cuidadoso* ‘careful, wary’ are proposed to be opposites to *dormido* ‘asleep’, *entregado al sueño* ‘asleep’ (Ex. (110)), and *desmayado* ‘fainted. Whereas *despierto* and the other types in its cluster portray a change of state when used in *quedar(se)* + ADJ in which the subject becomes more alert mentally, *dormido* and its companion types portray a loss of alertness. In Example (109), the subject undergoes a change of state in which he recovers both consciousness (i.e. *quedó despierto* ‘he woke up’) and his previous vigor despite suffering a blow that knocked him to the ground unconscious. An opposite change is seen in Example (110) in which the subject looses consciousness.

(109) *Hirióle el primer golpe en descubierto*
y dio con él en tierra sin sentido,
aunque con brevedad quedó despierto,
a su antiguo vigor restituýdo.
‘He wounded him with the first uncovered blow
and it threw him to the ground senseless,
although quickly he woke up,
to his previous vigor restored. (*Genealogía de la toledana disreta*. 17th c. (1604). E. Martínez. Lemir.)

(110) *Rindiose, pues, Felicio a los combates de sus pensamientos y, sin querer, quedó entregado al sueño.*
‘Felicio gave in, then, to the battles of his thoughts and, without intending, he was delivered into sleep.’ (*La Constante Amarilis.* 17th c. (1609). C. Suárez de Figueroa. Lemir)

5.3.1.6 The *alegre / satisfecho* 'happy / satisfied' clusters in the 1700's

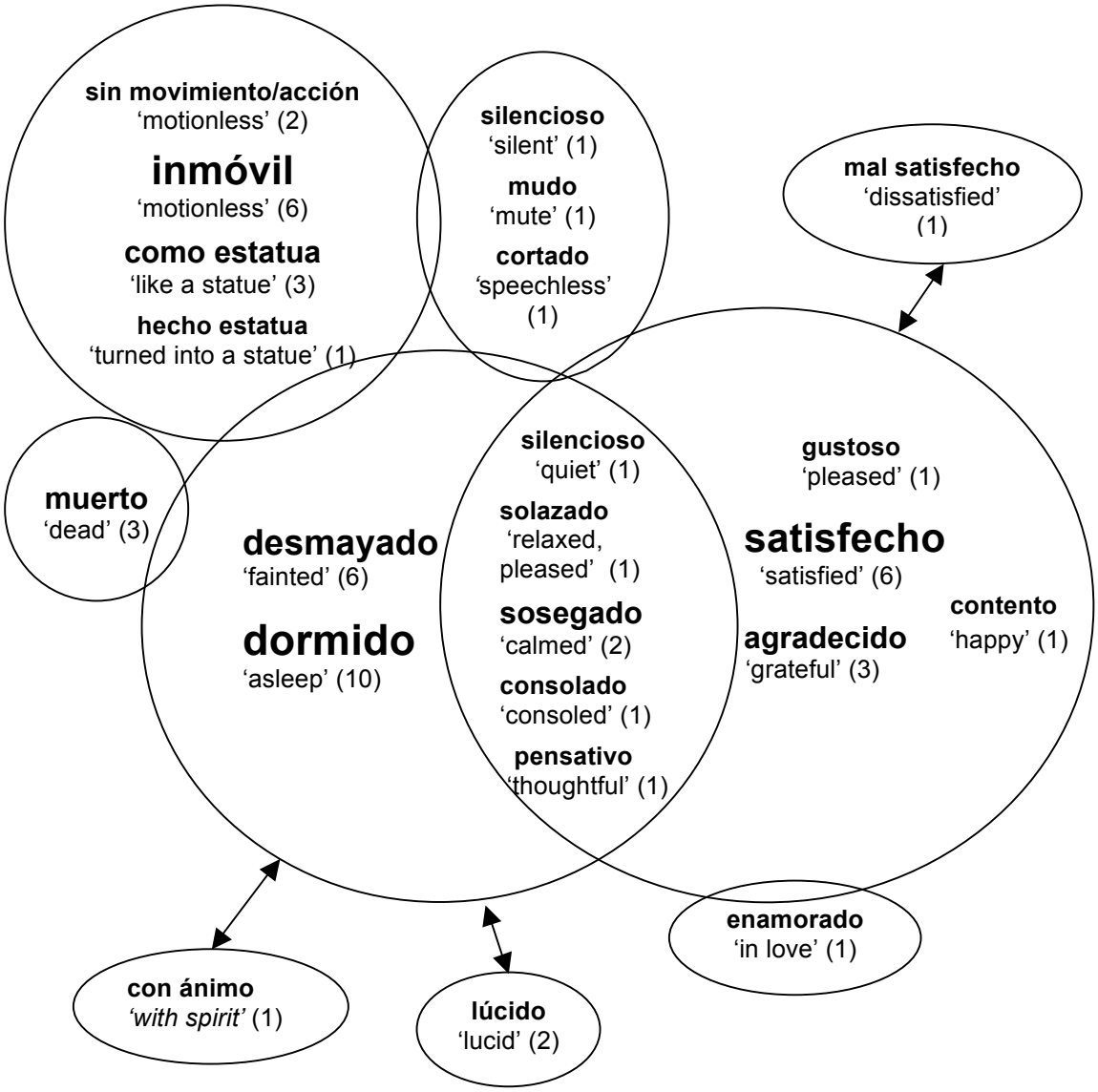


Figure 32, 1700's: The *alegre / satisfecho* clusters.

In the 1700's, the *alegre / satisfecho* clusters continue to maintain the same overall structure from previous centuries, as shown in Figure 32. Many of the established types show a rise in token frequency in this century. Appearing six times in the more modern form, *inmóvil* 'motionless' emerges with enough tokens to put it above the

proposed CIC threshold. *Desmayado* 'fainted', and *dormido* 'asleep' also appear with enough tokens (six and ten respectively) to surpass the CIC threshold.

Some established types showed a decline in token frequency in the data for this century. *Quedar(se) muerto* 'to be left dead' appears in the data with a token frequency that puts it below the CIC threshold for the first time since the 1300's. Likewise, *contento* 'happy' appears below the CIC threshold with only one token. The type *quieto* 'still, calm' did not appear in the data for this century but a similar type, *pensativo* 'thoughtful', emerges. It will go on to appear in the two following centuries.

With ten occurrences, *quedar(se) dormido* 'to fall asleep' has surpassed *satisfecho* 'satisfied', or any other types, in token frequency. Evidence that *quedar(se) dormido* 'to fall asleep' has increased in overall frequency is provided in this study and in the *Corpus del español* (Davies 2002-), as shown in Table 23, below. In both sources of data, *quedar(se) dormido* did not emerge until the 1500's.

Table 23. Overall frequency of *quedar(se) dormido* 'to fall asleep'.

	Data for present study: occurrences per 10k	*Corpus* (Davies 2002-): occurrences per million
1500's	0.02	0.8
1600's	0.02	1.2
1700's	0.22	1.4
1800's	0.08	3.4
1900's	0.28*	22[27]

Before *dormido* 'asleep' appeared in the data in the 1500's, three similar types occurred in the 1400's: *desmayado* 'fainted', *amortecido* 'fainted', and *cansado* 'tired'. None of these types had occurred previously in the data and their appearance coincides with the appearance of 'calm' types argued in Section 5.3.1.3 to relate to *alegre* (e.g.

[27] This figure come from the fictional register in Davies' (2002-) data.

sosegado ‘calm’, *asosegado* ‘calm’, and *confortado* ‘comforted’). This also coincides with the emergence of the *muerto* clusters. Because *muerto* ‘dead’ does appear above the CIC threshold in the 1500’s, the types *desmayado* ‘fainted’, and *amortecido* ‘fainted’ most likely emerged based on analogy to *muerto* ‘dead’. However, it is plausible that *cansado* ‘tired’, particularly, could have emerged based on *sosegado* ‘calm’, *asosegado* ‘calm’, and *confortado* ‘comforted’ since these denote a change of state in which the subject becomes less energetic and lively without being such a terminal state as *muerto*. Nevertheless, by the 1500’s *dormido* emerged in usage with *quedar(se)* + ADJ whether based on analogy to types such as *desmayado* ‘fainted’, and *muerto* ‘dead’, to types such as *sosegado* ‘calm’ and *alegre* ‘happy’, or to both.

Example (111) shows how *sosegado* can denote a change resulting not only in calmness but in a certain degree of satisfaction. The foreign captain and the other subjects became calm as a result of being satisfied with the offerings of the captain and the thanks offered by the people. The situation further improved when the surgeons informed the subjects that the injury wasn’t going to be grave; the subjects were consoled.

(111) *Con estos generosos ofrecimientos que les hizo el capitán extranjero, y que le agradecieron del mejor modo que les fue posible, quedaron sosegados todos, y con las nuevas que les dieron los cirujanos de que no era de cuidado la herida quedaron consolados.*
‘With those generous offerings that the foreign captain made them, and because they thanked him in the best way that they could, everyone became calm, and with the news that the surgeons gave them that the injury wasn’t grave, they were consoled.’ (*Los trabajos de Narciso y Filomela.*18th c. V, Martínez Colomer. BVMC.)

This is the last century in which any ‘love’ types appear; *enamorado* ‘in love’ has appeared in the data since the 1400’s. Example (112) describes how Acca did not fall in love with herself as Narcissus did.

(112) *Ni la sucia y asquerosa vieja Acca, cuando mirando en un espejo su disforme y feo semblante, quedó enamorada, ciega y celosa de sí misma.* 'Not even the dirty and disgusting old lady Acca, when looking at her deformed and ugly face in a mirror, was in love, blind, and jealous of herself.' (*Fray Gerundio de Campazas.* 18th c. (1758). J. F. de Isla. BVMC.)

5.3.1.7 The *alegre / satisfecho* 'happy / satisfied' clusters in the 1800's

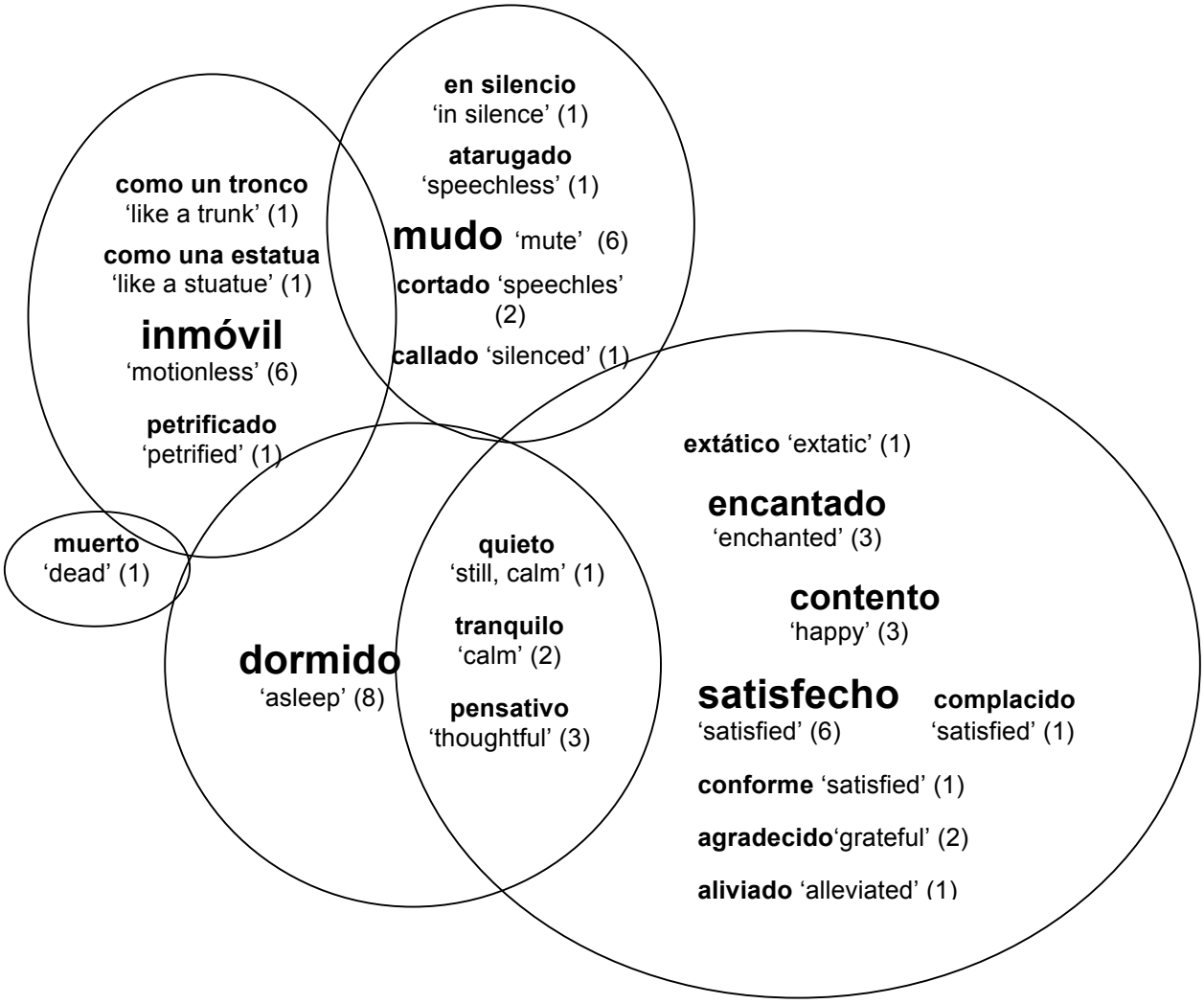


Figure 33, 1800's: The *alegre / satisfecho* clusters.

Even though, as seen in Figure 33, this set of clusters resembles the ongoing structure of these clusters since the 1400's, some key types have disappeared from the

data. *Desmayado* ‘fainted’, which first appeared in these data in the 1400’s, does not appear in the 1800’s. As mentioned above, *enamorado* ‘in love’, which also appeared from the 1400’s to the 1700’s, does not appear in this century either. *Dormido* ‘asleep’ continues to have a higher token frequency than *satisfecho* ‘satisfied’.

All three of the ‘calm’ types that appear in the data in this century (e.g. *quieto* ‘still, calm’, *tranquilo* ‘calm’, and *pensativo* ‘thoughtful’) will go on to appear in Bybee & Eddington’s (2006) data from the 1900’s. With 29 occurrences in the written data (but none in the spoken), *quedar(se) quieto* ‘to become still’ will be the most highly occurring verb + adjective combination in their data. *Quedar(se) tranquilo* ‘to become calm’ appears in their data at a relatively high token frequency as well with six occurrences in the spoken data and ten in the written data.

5.3.1.8 Overview of the *alegre / satisfecho* clusters

One of the main things that this set of clusters shows is that it is possible for seemingly unrelated adjectives to be linked by a series of intermediate ones. For example, the adjective *dormido* ‘asleep’ was initially proposed to belong to the *alegre* / *satisfecho* clusters based on the organization of a similar cluster in Bybee & Eddington (2006: Fig. 3, p. 334) centering on *quedarse quieto* ‘to become calm’. Their cluster organization was based on the placement of adjectives according to perceived similarity of a native speaker participant. Extrapolating from this set of clusters in Bybee & Eddington (2006), one could propose the following connection between an adjective such as *alegre* and *dormido* based on the following interposing central members found there: *contento* ‘happy’—*satisfecho* ‘satisfied’ —*tranquilo* ‘quiet, peaceful’ — *quieto* ‘calm’ — *callado* ‘quiet’ — *dormido* ‘asleep’. Furthermore, and adding one more link, the adjective *muerto* ‘dead’

was placed in close proximity to *dormido* 'asleep' by Bybee & Eddington's (2006) participant. However, because *quedar(se) muerto* 'to be left dead' is proposed to be the central member of a productive category with many types, it was also shown in its own set of clusters with all of the included types in Appendix 1, Section 5.

As mentioned above in Section 5.2.1.5, *inmóvil* is also related to types that portray changes of surprise or fear in the *confuso / suspenso* 'confused / surprised' clusters (see Appendix 1, Section 3). This chain of linked clusters also plausibly leads to the *confuso / suspenso* 'confused / surprised' clusters (Appendix 1, Section 3) in other ways. The main link is through the type *encantado* 'enchanted' which, as in English, can convey a reaction of being delighted by something or of being captivated to the point of astonishment. Taken together, all of these pathways of linkages shows that it is possible that all adjective types may be linked to each other through a series of intervening types. Using the exemplar model enables an analyst to recognize this; the clusters proposed here are not intended to represent discrete classes of adjectives that belong only to one category, but to represent the ways in which semantically close adjectives are related to each other.

This chain of related adjectives that contains unrelated types on either extreme was later supported by the similarity study (Ch. 6) conducted subsequent to the analysis. *Satisfecho* and *contento* were perceived in the aggregate as being similar. The adjective *tranquilo* 'quiet, peaceful' could be a linking member as it had a high average similarity rating to *contento* 'happy' and to *satisfecho* 'satisfied' on one side, and to *dormido* and to *quieto* 'calm' on the other. Following the chain, *quieto* had an very high average

similarity to *inmóvil* ‘motionless’, which was perceived as similar to *muerto* ‘dead’. See Chapter 6 for more details.

Concerning the extension of a category through a series of linked types, the *triste* clusters (shown in Appendix 1, Section 2) are proposed to be linked to the *alegre / satisfecho* clusters because they are opposites. They were not included in the analysis, but the reader should bear in mind that there are ‘happy’ and ‘sad’ adjectives used in *quedar(se)* + ADJ. Accordingly, there could be an analysis of these clusters exploring their relationship as opposites. This was not shown here because an analysis based on opposites was shown in the *pobre / rico* clusters in Section 5.2.3.

In terms of diachronic tendencies in categories, this set of clusters shows how the central member may change over time; it shifted from *alegre* ‘happy’ to *contento* ‘happy’ to *satisfecho* ‘satisfied’. Appearing in the data in the 1200’s, *alegre* was the first adjective that meant ‘happy’ that occurred in the data. By the 1400’s, *alegre* appeared with four tokens, enough to put it above the CIC threshold, and was surrounded by synonymous types (e.g. *contento* ‘happy’, *ufano* ‘cheerful’, *gozoso* ‘joyful’) and other related types (e.g. *sin queja* ‘without complaint’, *sin pena* ‘without grief’, *acabado* ‘satisfied’, and *limpio de odio* ‘free of hate’) including three ‘calm’ types (e.g. *sosegado* ‘calm’, *asosegado* ‘calm’, and *confortado* ‘comforted’). All of these related and synonymous types had only one occurrence whereas *alegre* had four, making it the clear choice of the central member. In the next century the central member was proposed to have shifted to *satisfecho* ‘satisfied’; despite the fact that it had one fewer token than *contento* ‘happy’, it is more general and had a notably higher standardized frequency in Davies’ (2002-) data. By the 1600’s with nine occurrences, *satisfecho* did surpass *contento* and became the

central member of the *alegre / satisfecho* clusters to the 1800's, a figure bolstered by a widening standardized frequency gap in Davies' corpus. Even though this category seems to have begun to fade, *satisfecho* is the proposed central member continues through the 1900's (Bybee & Eddington 2006). With three occurrences in the 1900's (two in the written data, one spoken in the spoken data), *satisfecho* continues to occur at a slightly higher rate than *contento*, which appears once in the spoken data.

Despite the changes in the central members of the *alegre / satisfecho* clusters over time, these clusters display a high degree of coherency in that the overall structure does not change. One of the main reasons that the structure endures is because the category is based on meaning. Table 24 gives a timeline of the types found in the *alegre / satisfecho* clusters including *mudo* 'mute', *inmóvil* 'motionless' and *muerto* 'dead'. The analysis of the *alegre / satisfecho* clusters shows how formulaic language has longevity. *Quedar(se) satisfecho* 'to become satisfied' appeared above the proposed CIC threshold for four centuries in a row before losing relative token frequency in the 1900's. The types *dormido* 'asleep', *inmóvil* 'motionless', and *muerto* 'dead' were CIC's in these data for three consecutive centuries. *Quedar(se) alegre* 'to become happy' was most likely a CIC in the 1500's even though it idiosyncratically did not appear in these data in that century. Yet, in the *Corpus del español* (Davies 2002-) this combination occurred six times in the 1500's indicating the possibility that it was a CIC.

Table 24. Types from the *alegre / satisfecho* clusters that appear in at least 3 centuries.

	1200	1300	1400	1500	1600	1700	1800	1900
alegre 'happy'	=====	====	####	DDD	####			
contento 'happy'			====	####	####	====	====	====
satisfecho 'satisfied'			DDD	####	####	####	####	====
ufano 'cheerful'			====	====	====	DDD	DDD	
enamorado 'in love'			====	====	====	====	DDD	
sosegado 'calm'			====	====	====	====		
pensativo 'thoughtful'				====		====	====	====
quieto 'still, calm'					====	DDD	====	####
dormido 'asleep'				====	====	####	####	####
desmayado 'fainted'			====	====	====	####	====	
mudo 'mute'			====	====	####	====	####	====
inmóvil 'motionless'					====	####	####	####
muerto 'dead'	====	DDD	####	####	####	====	====	====

Key: (====) = presence of type in data below CIC threshold
(####) = presence of type in data above CIC threshold
(DDD) = presence only in Davies *Corpus del español* (2002-)

These data also provide evidence that there is a pattern to the emergence of the different types and that frequency shifts are not random. In some cases, as is the case with *contento* 'happy', *satisfecho* 'satisfied', *desmayado* 'fainted', and *muerto* 'dead', the type emerges, gains in token frequency to the point where it surpasses the CIC threshold, and then diminishes in frequency. Therefore, even after loosing CIC status according to the proposed method of determination, a type may continue to be used for many centuries in *quedar(se)* + ADJ, as in the case of *contento* 'happy' and *muerto* 'dead'[28]. In other cases, as with the adjectives *quieto* 'calm, still', *dormido* 'asleep', and *inmóvil* 'motionless', the type was seen in the data and then underwent a rise in token frequency that has persisted until the 1900's, the last century from which data has been gathered. If other types serve

[28] This points to the notion that the CIC threshold may be too high and that there is evidence of conventionalization in types with lower token frequency. However, a conservative measure was applied to these data in order to exclude borderline cases, as was done by Erman & Warren (2000) in their study of prefabs.

as an example, these will begin to lose token frequency in written Spanish at some point in time.

There are only three types that do not appera in the data with a consistent pattern. *Alegre* 'happy' was not found in the data in the 1500's even though there were six occurrences in the *Corpus* (Davies 2002-). The fact that it does not appear in this century in my data seems idiosyncratic considering that it appears in the data above the CIC threshold in the two surrounding centuries. After appearing in the 1500's, *pensativo* 'thoughtful' did not occur in these data or in the *Corpus* (Davies 2002-) in the 1600's. However, it resurfaced in these data in the 1700's and persisted into the 1900's. *Mudo* 'mute' appeared in these data in the 1400's and rose to above the proposed CIC threshold in the 1600's, diminished in frequency in the 1700's, and rose back up to above the CIC threshold in the 1800's.

Despite these three anomalous patterns, it is common for types that increase in token frequency beyond the CIC threshold to exhibit a bell curve of frequency. As with *rico* 'rich', *pobre* 'poor', and *libre* 'free', the types *contento* 'happy', *satisfecho* 'satisfied', *desmayado* 'fainted', and *muerto* 'dead' were first found in the data at a low token frequency before gaining considerably in token frequency, and then diminishing in frequency.

5.4 Summary

This chapter has presented an analysis of four sets of proposed exemplar clusters of adjective types appearing in *quedar(se)* + ADJ. One of the main goals was to apply the exemplar model of categorization to diachronic data in order to account for changes in usage. Aside from emergent forms, analysis of the data revealed that there was also a

great deal of endurance over time; groups of related adjectives appear consistently throughout large spans of time despite the emergence of new forms and the loss of other ones. These groups of adjectives plausibly form gradient categories that are organized around a high frequency central member. Regarding individual types, not only do highly frequent combinations of *quedar(se)* + ADJ show longevity, but some lower frequency combinations endure diachronically as well. Over time, however, categories tend to mutate to varying degrees in terms of the types involved and the token frequency of these types.

Some sets of exemplar clusters, such as the *solo* clusters (Section 5.2.1) show that it is possible for a category to be stable over time and mutate very little; the central member never changes and many types appear over the span of several centuries. One of the most significant changes in these clusters was the fact that the central member, *quedar(se) solo* 'to be left alone', gained precipitously in token frequency in the 1800's. Because *quedar(se) solo* appeared above the CIC threshold in every century from which data was collected, and above the prefab threshold from the1500's onward this set of clusters shows that prefabs have longevity.

The *libre* clusters (Section 5.2.2) showed increasing cohesion over time. By the 1400's *quedar(se) libre* 'to be set free' emerged as the central member and maintained that status through time despite fluctuations in the category. *Quedar(se) libre* underwent an increase in overall frequency before losing frequency that seemed to correspond with a gain and decline in type frequency. In these data the overall frequency of *quedar(se) libre* peaked in the 1500's, and in the *Corpus del español* (Davies 2002) it peaked in the 1600's. Although it appears in the *Corpus del español* (Davies 2002) in the 1900's, it

does not appear in Bybee & Eddington's (2006) data. Other types showed persistence in this set of clusters with a gain and loss of token frequency (e.g. *obligado* 'obligated', *preso* 'imprisoned'). Furthermore, *quedar(se) libre* appeared above the prefab threshold from the1300's to the 1900's, demonstrating that prefabs show endurance.

Other clusters have been less consistent through the years. One of the tendencies documented in the *rico / pobre* clusters (Section 5.2.3) was that the two clusters analyzed, related through being opposites, were less coherent than the *solo* clusters. As the categories emerged, the proposed structure changed from one century to the next, possibly due to the fact that the central members (*quedar(se) rico* and *quedar(se) pobre*) were less frequent in the data relative to more robust ones (e.g. *quedar(se) solo, quedar(se) libre*). The 1400's and the 1600's were the only centuries in which any type appeared above the CIC threshold in these clusters. This set of clusters reaches maximum expansion in type frequency in the 1400's and then begins to diminish. By the 1900's this set of clusters seems to have become unproductive as none of the types found in these data appeared with *quedar(se)* + ADJ in Bybee & Eddington (2006). However, they found occurrences of *rico* 'rich' with the verb *hacerse*. Despite the fact that they were less coherent, these clusters demonstrate how the current analysis can be applied to categories that are related through having opposite central members. In every century, there were types that showed a gain in wealth and types that showed a loss of wealth.

It is possible for groups of seemingly unrelated adjectives to be linked through intermediary clusters. This was demonstrated in the *alegre / satisfecho* clusters and has lead to the observation that all clusters could somehow be linked. The proposed structure was supported by the similarity study in the following chapter. These clusters underwent

a noticeable expansion in the 1400's. The structure of these clusters remained stable over time even though the central member of the main cluster, the *alegre* cluster, changed from *alegre* 'happy' to *contento* 'happy' to *satisfecho* 'satisfied' over time. Similar types went on to occur in the 1900's in Bybee & Eddington's (2006) data.

There was a tendency for the emergence of high frequency central members to correspond to an increase in type frequency of related types. Although it is a phenomenon that has occurred in other clusters, it was discussed in most detail in the *solo* clusters (Section 5.2.1) and the *libre* clusters (Section 5.2.2). One proposed reason for this correlation is that, as a type is used more frequently, it has more opportunities to be a model of analogical extension. This seems to be the case with the *solo* clusters. However, with the *libre* clusters, the increase in the token frequency of *quedar(se) libre* is preceded by a gain in type frequency of this category. Nevertheless, as *quedar(se) libre* loses overall frequency, the category diminishes in type frequency.

In order to analyze formulaic usages of *quedar(se)* + ADJ, a method of identifying both conventionalized instances of constructions (CIC's) and prefabs was proposed in Chapter 4:, Section 4.2.1.1. The main distinctions between the identification of the two types of formulaic sequences are that the criteria for CIC's are derived from these data, are calculated using one frequency measurement (twice the token / type ratio), and is used on a broader group of types. On the other hand, the prefab threshold takes into account two measurements of relative frequency, uses the *Corpus del español* (Davies 2002-), and is applied only to the central category members *quedar(se) solo* and *quedar(se) libre*. This was useful in supporting the observation that prefabs have longevity and that formulaic sequences attract new types through analogical extension.

Also, it gives an idea as to the distribution of formulaic versus novel forms. Taking into account all the data from seven centuries, 28.5% of all types were novel, only occurring once in all of the data. These novel types only accounted for 16% of all tokens. On the other hand, 10.5% of all types found in these data appeared above the CIC threshold and these accounted for 34% of all tokens. Even though novel forms have a higher type frequency, they account for fewer tokens than conventionalized sequences in these data as a whole. The higher token frequency of formulaic sequences is also observed in Erman & Warren (2000) who found that at least 55% of their written data consisted of prefabs. However, the distribution of CIC's versus novel usages changed over time. CIC's appeared at the highest ratio in the 1600's accounting for 14.5% of all types and 43% of all tokens. Over the span of time studied, novel types emerged as accounting for a greater portion of the data than CIC's. By the 1800's novel types accounted for 44% of all types and 24.5% of all tokens. In this same century CIC's accounted for only 7% of all types and 32% of all tokens. It appears that as the overall construction becomes less frequent from the 1600's to the 1800's, that usage becomes more scattered among novel types instead of being concentrated in types with higher token frequency, despite the fact that some types continue to rise in overall frequency (e.g. *quedar(se) solo* 'to be left alone').

Even though the present method of analysis has revealed many diachronic trends regarding semantic categories of adjectives of the construction *quedar(se)* + ADJ, it is not without limitations. One of the limitations stems from idiosyncrasies in the data obtained in which key types failed to appear in centuries in which their appearance was anticipated. These unexpected absences were compensated for by coordination with data from the *Corpus del español* (Davies 2002-). Another limitation is in the model of

representation. It would be more revealing to portray these categories in a three dimensional model that showed how there may be even more connections between clusters than is proposed here.

Chapter 6: Similarity Experiment

The structure of the clusters analyzed in Chapter 5 was largely confirmed by a similarity experiment presented in this chapter. Using a multidimensional scaling (MDS), a multivariate analysis was performed using data from questionnaires filled out by native speakers of Spanish. The questionnaires gave the participants the task of rating the similarity of pairs of adjectives on a scale of one to five. Despite a few differences, this similarity study resembles the one reported in Bybee & Eddington (2006: 347-9). An MDS analysis is an supplement for the present study; it generates maps of perceptual space that resemble the configurations of perceptual space presented in the exemplar clusters in the previous chapter. The MDS analysis provides a data-driven spatial analysis that could be compared to the theoretical clusters proposed in the previous chapter. While the perceptual maps generated by MDS analysis may resemble semantic maps used by typologists, the key difference lies in the fact that MDS is mathematically and computationally formalized whereas semantic maps are created through the intuition of the analyst and are not formalized (Croft & Poole 2008).

6.1 Data: the questionnaires

There were two questionnaires administered to participants: Questionnaire A had all possible pairs of 19 different adjectives, resulting in a total of 171 test items, and Questionnaire B had 20 different adjectives and 190 test items. As with the one used in Bybee & Eddington (2006) the participants were asked to rate the similarity of adjectives on a five point scale with the number one corresponding to *nada parecido* 'not similar', three corresponding to *algo parecido* 'somewhat similar', and five corresponding to *muy parecido* 'very similar'. There was no context given with the adjectives on the

questionnaires. Adjectives came mainly from the data from the 1800's with some coming from earlier centuries. The questionnaires did not ask for any demographic information except that the participant acknowledge that they were a native speaker of Spanish.

The adjectives from Questionnaire A came from three sets of clusters that describe social states (e.g. the *solo* 'alone' clusters, the *rico / pobre* 'rich / poor' clusters, and the *libre* 'free' clusters) based on the presumption that these will have many overlapping members. A list of all of the adjectives in this questionnaire is provided in Table 25. The objective of submitting these pairings to an MDS analysis was to determine whether native speakers would rate the adjectives from each cluster as more similar to one another than to adjectives from different clusters. In other words, would the MDS analysis confirm the proposed exemplar clusters?

Table 25. Adjectives on Questionnaire A (N=19).

solo 'alone' clusters	*rico / pobre* 'rich / poor' clusters	*libre* 'free' clusters
solo 'alone' *sin padre* 'without father' *a solas* 'alone' *viudo* 'widowed' *huérfano* 'orphaned' *perdido* 'lost' *desamparado* 'separated, unprotected' *con hijo* 'with a child'	*pobre* 'poor' *rico* 'rich' *con galardón* 'with a reward' *sin dinero* 'without money' *heredero* 'heir' *con honra* 'with honor'	*libre* 'free' *perdonado* 'pardoned' *absuelto* 'absolved' *preso* 'imprisoned' *obligado* 'obligated'

The adjectives used in Questionnaire B came from two sets of clusters depicting emotional states (e.g. the *alegre / satisfecho* 'happy / satisfied' clusters, and the *confuso / suspenso* 'confused / surprised' clusters), also based on the presumption that these would have many overlapping members. A list of the adjectives is provided in Table 26. Some of the adjectives appeared in both of these sets of proposed exemplar clusters (e.g. *mudo* 'mute', and *inmóvil* 'motionless'), but were deemed to be more related to the *confuso /*

suspenso clusters in the context of tokens of *quedar(se)* + ADJ. Likewise, part of the purpose in choosing these adjectives was to test whether speakers perceived *mudo* and *inmóvil* as being more related to adjectives denoting calmness (e.g. *tranquilo* 'calm', *quieto* 'calm, still'), physical states in which there is no movement or speech (e.g. *muerto* 'dead', *dormido* 'asleep'), or to that portray states of fear or surprise (e.g. *espantado* 'scared', *sorprendido* 'surprised'). A second goal, as with the previous questionnaire, was to determine if speakers rated the types in each proposed cluster as being similar to one another therby confirming the structure of the proposed clusters.

Table 26. Adjectives on Questionnaire B (N=20).

alegre / satisfecho 'happy / satisfied' clusters	*confuso / suspenso* 'confused / surprised' clusters
satisfecho 'satisfied'	*mudo* 'mute'
tranquilo 'calm'	*inmóvil* 'motionless'
pensativo 'thoughtful, calm'	*confuso* 'confused'
quieto 'still, calm'	*suspenso* 'surprised'
contento 'happy'	*sorprendido* 'surprised'
encantado 'enchanted'	*estupefacto* 'stupefied'
dormido 'asleep'	*asombrado* 'astonished'
muerto 'dead'	*maravillado* 'astonished'
	atónito 'astounded'
	petrificado 'petrified'
	espantado 'scared'
	aterrado 'scared'

6.2 Participants

Participants were recruited from the Department of Spanish and Portuguese at the University of New Mexico in the Spring semester of 2009. The questionnaire was given to native Spanish speaking faculty and to graduate student TA's that volunteered to participate. Several TA's recruited native Spanish speaking undergraduate students from Upper Division Spanish classes as an opportunity to earn extra credit. The TA's that

recruited undergraduate students were doctoral students in the Linguistics program and taught classes such as Phonetics, Advanced Grammar, and Introduction to Linguistics. A total of 31 participants completed Questionnaire A, and a total of 33 different participants completed Questionnaire B.

6.3 Methods

The ratings from the questionnaires was averaged and formatted as a similarity matrix showing the ratings for each pair. This matrix was submitted to statistical program, PerMap (Heady 2007), designed specifically for carrying out an MDS analysis. Since MDS is more common in fields other than linguistics, such as political science, sociology, psychology, and marketing (Stalans 1995), a brief explanation is necessary.

6.3.1 MDS, a brief explanation

MDS is a pairwise method of multivariate analysis that represents measurements of similarity (or dissimilarity) of objects in a category as distances between points of multidimensional space, most commonly on two or three dimensions (Borg, & Groenen 1997). The purpose of representing data in such a way is to reveal patterns in similarity data that might not be otherwise accessible (Stalans 1995). This procedure is referred to by Skupin & Butterfield (1997: 117) as *spacialization* which they define as "a projection of elements of a high-dimensional information space into a low-dimensional, potentially experiential, representational space". This allows the researcher to explore what features may underlie judgments in similarity among members of a category by applying their understanding of the different aspects of the items to the analysis (Stalans 1995). Because they are meant to capture participant perceptions, the spatial plots created by MDS analyses are referred to by terms such as perceptual maps, sociograms, sociometric maps,

relationship maps, and concept maps (Heady & Lucas 2007). Stalans (1995) refers to the results of MDS analysis as a configuration. The visual plots created by the MDS analysis will be referred to in this study as both perceptual maps and configurations.

In order to explain perceptual maps created through MDS analysis, Stalans (1995), Croft & Poole (2008), and Bybee & Eddington (2006) used the analogy of constructing a distance map of major cities in which distance was calculated in terms of airline mileage. A table could be created in which the distance between each pair of cities was displayed, as is done in many atlases. Because it resembles a similarity matrix, the table containing the distances could be subjected to an MDS analysis. The map produced through calculated distances would create a configuration of the cities that roughly resembled a geographic map.

In analyzing a perceptual map, because it provides a visual representation of the patterns of perceived relatedness among items, researchers must identify relevant dimensions (Stalans 1995, Borg, & Groenen 1997). Often, the items in the perceptual map can be divided according to an axis that is applied by the researcher to the map. In the analogy of the distance map, above, a north / south axis, and an east / west axis could be drawn on the distance map created by the MDS analysis (Stalans 1995). In the perceptual maps created by this analysis, axes are be applied to clusters of adjectives that divide them along dimensions that largely correspond with the categories proposed in the previous chapter.

Borg & Groenen (1997: 3) propose that there are four purposes for conducting an MDS analysis: (a) MDS as a method of analysis that presents similarity data as distances in a low-dimensional plot that can be visually inspected, (b) a method of providing

empirical evidence to support whether or not criteria proposed by researchers for distinguishing between items is mirrored in participant perception, (c) a method of data analysis that allows the researcher to discover the factors that underlie judgments of (dis)similarity, and (d) a method of creating a psychological model that explains dissimilarity judgments in terms of a rule that portrays a certain type of distance function.

Since much of the analysis in Chapter 5 was based on visual figures that represented categories of adjectives that appeared in *quedar(se)* + ADJ in a hypothetical psychological space (i.e. semantic maps), it is useful to have the configurations of an MDS analysis as a tool comparison. The perceptual maps generated in this analysis supported the construction of the categories presented in Chapter 5.

6.3.2 MDS, data & analysis

Using the averaged similarity values from the questionnaires, similarity matrixes were created for use as a data source for PerMap (Heady 2007). These are shown below in Tables 27 and 28. The similarity matrixes show the average ratings of similarity between pairs. In the previous chapter, it was useful to look at individual pairs in order to support some of the proposed clusters. However, as pointed out by Stalans (1995), the purpose of MDS is to provide researchers with a method of discovering patterns in data that might not be evident when examining raw data such as a similarity matrix. In the following section (6.4), many references will be made to these specific ratings from the matrices. The ratings were on a scale of one to five with one being the least similar and five being the most similar. Similar items are grouped close together and dissimilar ones far apart based on the aggregate ratings of their similarity.

The highest average similarity rating in Table 27 was 4.6 between the pair *pobre* 'poor' / *sin dinero* 'without money' and the lowest was 1.4 between a number of pairs which will be mentioned in the discussion of opposites below in Section 6.4.1.

Table 27. Similarity matrix for Questionnaire A.

	solo	sin padre	a solas	viudo	huerfano	perdido	desamparado	pobre	rico	con galardon	con honra	con hijo	sin dinereo	heredero	libre	perdonado	absuelto	preso
sin padre	3.57																	
a solas	4.1	2.8																
viudo	3.69	1.6	3.04															
huerfano	3.3	4.45	2.9	1.9														
perdido	2.9	1.9	2.98	2.3	2.5													
desamparado	3.48	3.75	3.5	2.7	3.9	3.04												
pobre	2.77	2.03	2.4	2	2.9	2.5	3.25											
rico	1.75	1.8	2.07	2.4	1.8	1.6	1.5	1.6										
con galardon	1.38	1.63	1.4	1.7	1.5	1.64	1.6	1.5	2.5									
con honra	1.4	1.6	1.6	2.2	1.7	1.75	1.5	2.5	2.55	3.4								
con hijo	1.68	1.67	1.5	2.2	1.7	1.55	1.5	1.7	2.26	2.15	2.8							
sin dinero	2.5	2.24	2.6	2.1	2.3	2.7	3.3	4.6	1.5	1.5	2.25	2.2						
heredero	2.2	2.6	2.16	2.3	2.5	1.6	1.85	1.7	3.6	2.37	2.8	2.27	1.46					
libre	2.8	1.9	2.2	3	2	1.7	2.1	1.9	3.1	2.04	2.7	1.6	1.89	2				
perdonado	2.4	1.6	1.4	1.6	1.5	1.7	1.7	1.6	1.6	1.92	2.3	1.4	1.6	1.6	2.9			
absuelto	1.96	1.57	1.8	1.86	2	1.96	1.8	2.3	1.9	1.59	2.35	1.65	2.07	1.7	2.7	2.9		
preso	2.43	1.7	2.6	1.5	1.8	2.07	2.17	2.25	1.7	1.36	1.4	1.6	2.35	1.85	1.73	1.65	2.1	
obligado	1.52	1.8	1.75	1.75	1.96	1.5	1.6	1.7	1.8	1.65	2.04	2.8	1.48	1.85	1.7	1.64	1.7	2.7

Table 28. Similarity matrix for Questionnaire B.

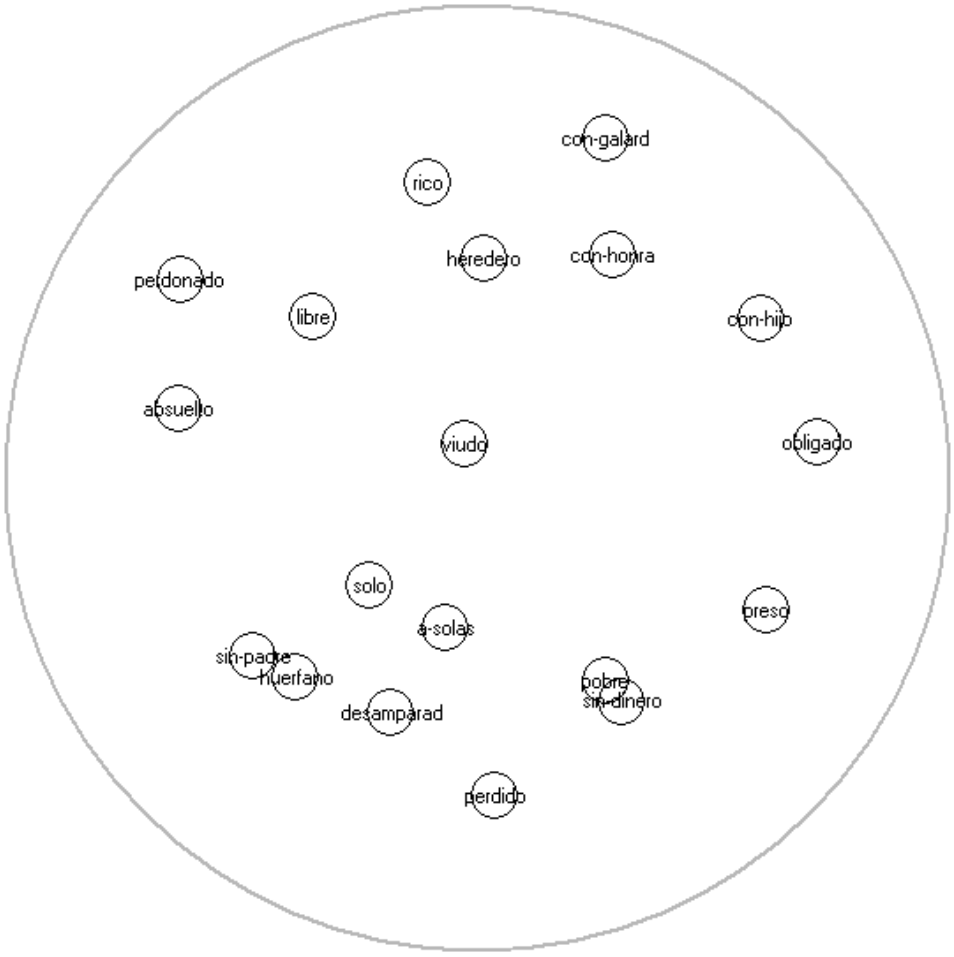
	satisfecho	tranquilo	pensativo	quieto	dormido	muerto	mudo	inmóvil	contento	encantado	suspenso	sorprendido	estupefacto	asombrado	confuso	aterrado	maravillado	atónito	espantado
satisfecho																			
tranquilo	2.6																		
pensativo	1.4	2.86																	
quieto	2.3	4.47	2.7																
dormido	1.9	3.03	1.5	3.2															
muerto	1.4	1.97	1.36	2.2	2.16														
mudo	1.36	2.3	2.1	2.9	1.86	2.4													
inmóvil	1.43	2.6	2.2	3.97	3.03	3.5	2.28												
contento	3.8	2.47	1.6	2.2	1.8	1.36	1.36	1.7											
encantado	3.15	2.3	2.03	1.6	1.8	1.1	1.4	1.6	3.97										
suspenso	1.47	1.97	2.1	2.09	1.9	1.6	2	2.7	1.3	1.34									
sorprendido	1.65	2.38	1.6	1.8	1.1	1.47	2	2.1	1.97	2.6	2.6								
estupefacto	2.15	1.8	2.03	2.4	1.6	1.3	2.47	2.7	1.9	1.86	2.64	3.2							
asombrado	1.78	1.9	1.8	1.8	1.47	1.6	2.78	2.6	2.2	4.2	2.47	4.8	2.4						
confuso	1.47	1.7	2.5	2.4	1.4	1.8	2.1	2.2	1.47	1.3	2.5	2.5	3.14	2.15					
aterrado	1.72	1.5	1.6	1.8	1.47	1.78	2.47	2.6	1.66	1.6	3	2.9	2.7	3	2.2				
maravillado	2.3	1.8	1.8	1.97	1.3	1.3	2.3	2.34	3.34	4.2	1.6	3.6	2.53	3.86	1.6	1.3			
atónito	1.8	2.03	1.7	2.2	1.4	1.9	2.6	2.97	1.79	2.26	2.5	2.97	2.68	3.16	2.16	2.4	3.2		
espantado	1.7	1.5	1.5	2.03	1.3	1.7	3.1	3.03	1.4	1.8	2.9	3.09	3.09	2.8	2.2	3.8	1.4	2.3	
petrificado	1.4	1.53	1.97	2.9	1.66	2.3	2.5	3.1	1.34	2.4	2.96	2.47	2.65	2.5	2.5	3.3	1.97	2.8	3.8

6.4 Results: Perceptual maps of Questionnaires A & B.

The figures in this section represent the configurations generated by PerMap (Heady 2007). The perceptual map generated through Questionnaire A is presented and analyzed in Section 6.4.1 and the one generated through Questionnaire B is presented and analyzed in Section 6.4.2.

6.4.1 Perceptual maps of Questionnaire A.

Figure 34. Perceptual map of Questionnaire A.



Looking at Figure 34, we see that the perceptual map generated by the MDS analysis has placed adjectives proposed to be related to one another in Chapter 5: (Sect. 5.2.1) as closer together while separating the ones that were not proposed to be related. The adjectives proposed to belong to the *solo* clusters (e.g. *solo* 'alone', *a solas* 'alone', *sin padre* 'without a father', *huérfano* 'orphaned', *desamparado* 'unprotected, separated', *perdido* 'lost', and *viudo* 'widow(er)ed') appear in close proximity and are not separated by any intervening adjectives. The adjective *viudo* 'widow(er)ed' presents a case worthy

of comment. While the similarity matrix (Table 27, above) shows that *viudo* is most closely related to *solo* 'alone' (3.69) and to other members of the *solo* clusters (*a solas* 'alone' (3.04), *huérfano* 'orphaned' (2.9), and *perdido* 'lost' (2.98)) it was also perceived as being related to *libre* 'free' (3.0). This could be due to the fact that participants equate marriage with a loss of freedom; by becoming widowed, a person gains freedom.

One of the opposites included in the *solo* clusters, *con hijo* 'with a child', showed some unexpected, but logical, connections. The fact that it was placed across the map from *solo* 'alone' is partially attributed to the fact that the two were not perceived as similar (1.68). However, it was also placed in the position that it is in because it was perceived as being most highly similar to *obligado* 'obligated' and *con honra* 'with honor' (both rated at 2.8). Just as participants may equate widowhood with freedom, they may equate having children with obligation. It is likely that *con honra* and *con hijo* may have been perceived as being more similar based on form. All of the items coming from *quedar(se) con* + NOUN (e.g. *con honra* 'with honor', *con hijo* 'with a child', and *con galardón* 'with reward') were grouped near one another presumably because of their structural similarity.

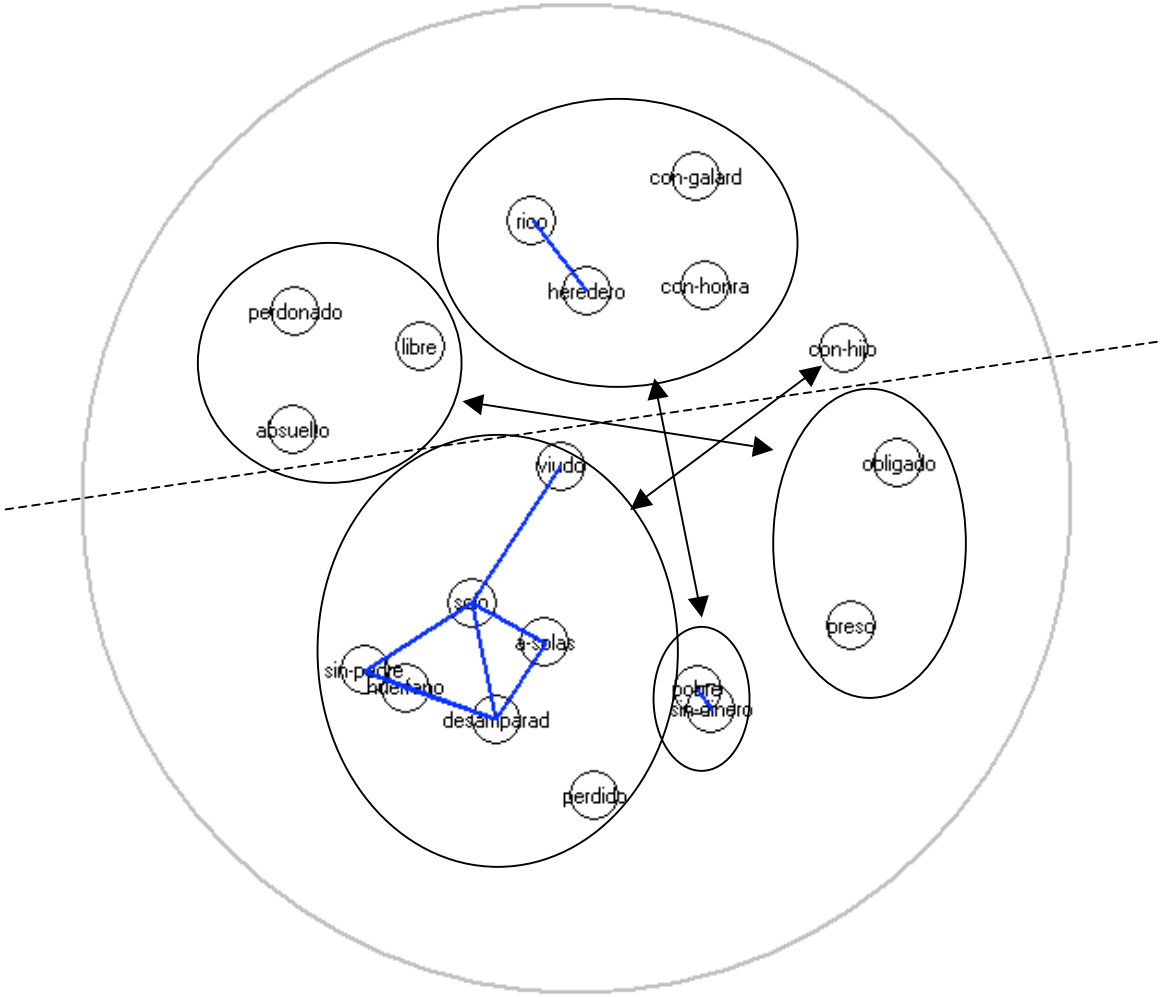
Adjectives belonging to the *libre* clusters (Chapter 5:, Sect 5.2.2) were grouped in a way that reflected the proposed clusters; *libre* 'free', *perdonado* 'pardoned', and *absuelto* 'absolved' were grouped close together. The proposed opposites, *obligado* 'obligated' and *preso* 'imprisoned', were close to one another but across the perceptual map from *libre*.

Finally *rico* 'rich' was placed near members proposed to be in its category (e.g. *con galardón* 'with reward', and *heredero* 'heir'). The opposites *sin dinero* 'without

money', and *pobre* 'poor' were placed across the perceptual map from *rico*. These latter types overlapped in the configuration due to the fact that this pair had the highest average similarity rating of all pairs (4.6).

It is also useful to look at the Waern Links when using the perceptual maps generated by MDS analysis in order to support the proposed exemplar clusters of the previous chapter. This function displays links between the most similar or dissimilar types according to a percentage specified by the analyst. Figure 35 includes the Waern Links of the 6% most similar links, represented by solid lines. The figure of 6% was chosen arbitrarily because it provides an idea of what items are similar without overcrowding the map. As mentioned previously in this chapter, researchers must identify relevant dimensions in analyzing a perceptual map by utilizing their knowledge of the items analyzed (Stalans 1995, Borg, & Groenen 1997). Consequently, Figure 35 includes some notation borrowed from the figures of the proposed exemplar clusters from Chapter 5. 'Bubbles' encircle groups of adjectives proposed to belong to exemplar categories based on semantic similarity. Arrows indicate an opposite relationship.

Figure 35. Perceptual map of Questionnaire A; 6% smallest links and proposed clusters.



As previously stated, the adjectives were grouped in a way that placed members of the proposed categories from Chapter 5: in proximity to one another. Because of this, it was possible to encircle types belonging to the same proposed categories; this is taken as evidence in support of the proposed structures in the previous chapter. However, there are some differences between the categories shown in 'bubbles' here and those from Chapter 5:. The type *con honra* 'with honor' was not proposed to be a member of the *rico / pobre* clusters through semantic similarity, but was placed in these to show structural similarity. Presumably because of this structural similarity, it was perceived as being

similar both to members of *quedar(se) con* + NOUN (e.g. *con galardón* 'with a reward' and *con hijo* 'with a child'), and to *rico* (2.5) and *heredero* (2.8). Nevertheless, the greater tendency is for adjectives from the proposed clusters to be grouped together, providing evidence that the proposed clusters in Chapter 5: depict a plausible representation of categories of adjectives found in *quedar(se)* + ADJ that a native Spanish speaker could have acquired through experience and usage.

The adjectives coming from the *solo* clusters show a great deal of cohesiveness as eight out of ten of the 6%[29] most similar links are between members of this cluster. Even though the structure of the clusters does not permit it to be clearly viewed, there are links between *sin padre* 'without father' and *huérfano* 'orphaned', between *huérfano* and *desamparado* 'unprotected, separated', and between *desamparado* and *sin padre*. The fact that *solo* 'alone' is linked to more types than any other provides evidence of its status as a central member of the category. In Chapter 5:, I claimed that *solo* was the most general adjective of the category; this could be part of the reason that it is linked to the most types. The only member that was not linked in the top 6% was *perdido* 'lost' which is, nevertheless, located close to the other members of the *solo* clusters.

Rico 'rich' and *heredero* 'heir' show a link in the top 6% as do *pobre* 'poor' and *sin dinero* 'without money'. One of the factors influencing the placement of *rico* in this configuration is its relative similarity to all of the *quedar(se) con* + NOUN types (e.g. *con honra* 'with honor' (2.55), *con hijo* 'with a child' (2.25), and *con galardón* 'with reward' (2.5)) and to *libre* 'free' (3.1). *Pobre* and *sin dinero* are located across the map from *rico* because of dissimilarity to this type (e.g. *rico / pobre* (1.6), *rico / sin dinero* (1.5)) and

[29] This figure (6%) was chosen because it cohesively showed similarities between types without overcrowding the configuration. In order to be consistent, it was applied to Figure 37 as well.

because they are similar to *desamparado* ‘unprotected, separated’ (3.25 and 3.3 respectively) and *preso* ‘imprisoned’ (2.25 and 2.35 respectively).

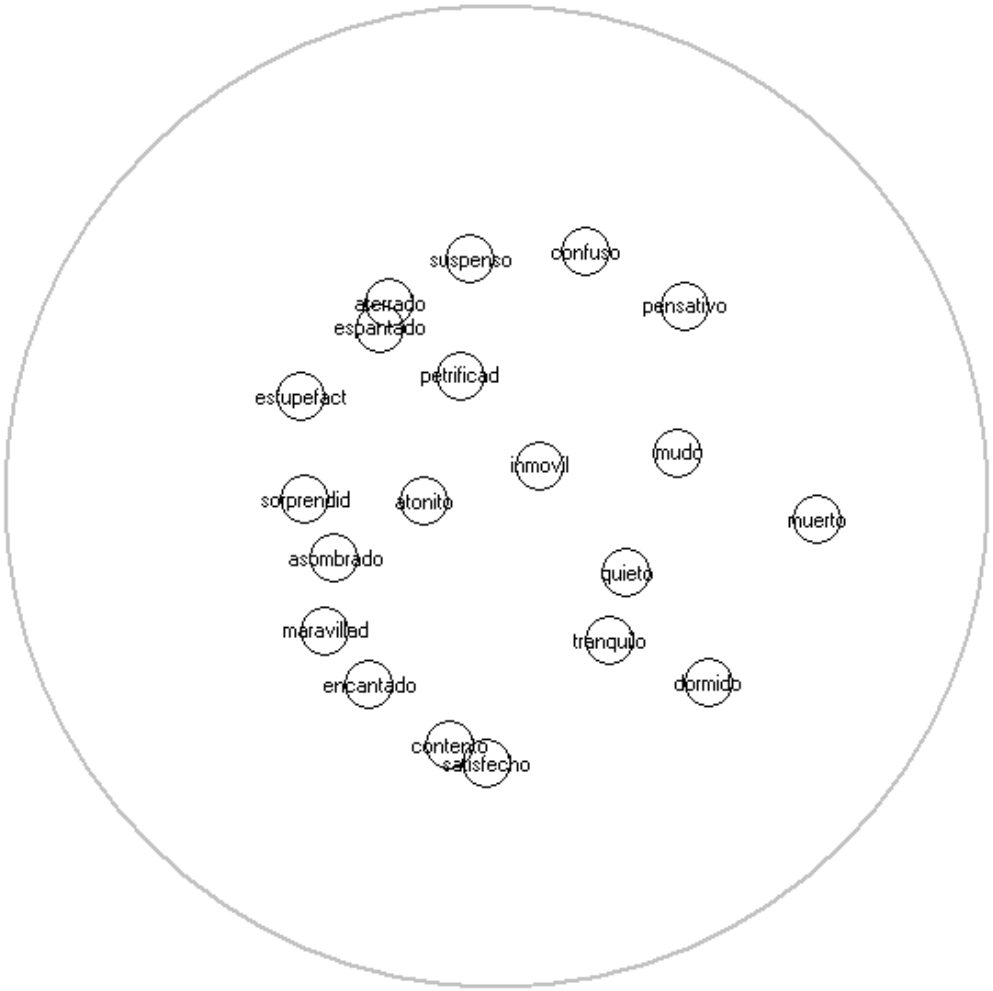
This distribution of types in the perceptual map draws attention to a pattern in which adjectives that convey a more positive, desirable state are in the top half of Figure 35 and the ones that convey more negative, undesirable state are in the bottom half. This ‘positive / negative’ division is represented by the dotted line. All of the arrows indicating opposites cross this dotted ‘positive / negative’ axis. While the types included in the *rico* cluster from Figure 35 portray a gain in material wealth (or honor in the case of *con honra*), the opposite types, *pobre* ‘poor’ and *sin dinero* ‘without money’ are located across the perceptual map and portray a loss of wealth. It is a similar configuration with the *obligado* and the *libre* clusters in Figure 35; they are proposed opposites (i.e. free vs. obligated or imprisoned), are across the map, and are on different sides of the ‘positive / negative’ axis. The third opposite relationship is proposed to be between the type *con hijo* ‘with child’ and the types in the *solo* cluster; the *solo* types depict states in which the subject is without the company of members of their social group. The type *con hijo* depicts an opposite state in which the subject is in the company of another human. While this pattern of was unexpexted, it reveals that speakers may perceive similarity between seemingly unrelated adjectives based on whether or not they represent positive or negative states. It would be informative to expand upon this in future studies.

It is worthwhile to note one final tendency regarding proposed opposites. In these data these pairs do not have the lowest similarity ratings. As mentioned above, the lowest average rating found in the similarity matrix in Table 27 was (1.4). There were five such

pairs: (a) *con honra* / *solo* ‘with honor / alone’, (b) *con galardón* / *a solas* ‘with reward / alone’, (c) *perdonado* / *a solas* ‘pardoned / alone’, (d) *preso* / *con honra* ‘imprisoned / with honor’, and (e) *perdonado* / *con hijo* ‘pardoned / with a child’. None of these least similar pairings include proposed opposites. Instead, they most likely have such low ratings because they are perceived to have little in common semantically. Even though it was by a small margin, the proposed opposites had slightly higher similarity ratings than these semantically unrelated types: *rico* / *pobre* ‘rich / poor’ (1.6), *libre* / *preso* ‘free / imprisoned’ (1.73), *libre* / *obligado* ‘free / obligated’ (1.7), *solo* / *con hijo* ‘alone / with a child’ (1.68). This is evidence that, while opposites are perceived as dissimilar, they are more similar than types for which no plausible semantic connection could be posited.

6.4.2 Perceptual maps of Questionnaire B.

Figure 36. Perceptual map of Questionnaire B.



As with the previous perceptual maps, Figure 36 supports the proposed category structure of the *alegre / satisfecho* clusters in (Chapter 5:, Sect. 5.3.1) in which intermediary types link 'happy' adjectives (e.g. *contento, satisfecho*) to types such as *dormido* 'asleep', and *muerto* 'dead'. The clearest intervening links here are *tranquilo* 'calm' and *quieto* 'still, calm', as proposed in Chapter 5:. Other links that do not intervene in a direct line in the figure, but play a role in this structure are *dormido* 'asleep', *inmóvil* 'motionless', and *mudo* 'mute'. Also, an important finding is that no

types from the *confuso / suspenso* clusters intervene in the set of adjectives portrayed in the *alegre / satisfecho* clusters. At the same time, these clusters are shown to be linked through *encantado* 'charmed' (as was proposed in Section 5.3.1.8).

In Chapter 5:, *inmóvil* 'motionless' and *mudo* 'mute' were proposed to be related to types denoting fear or surprise (e.g. *espantado* 'scared', *atónito* 'astonished'), to types denoting tranquility (e.g. *tranquilo* 'calm', *quieto* 'still, calm'), and to states that necessarily indicate physical inactivity such as *dormido* 'asleep' and *muerto* 'dead'. The MDS analysis supports this proposition. This series of links was overlooked by Bybee & Eddington (2006: 334) who proposed that it was more related to *quieto* 'calm, still' than to types that denoted fear or surprise. However, their assertion was not entirely without reason. As seen in Table 28, *inmóvil* was rated in the aggregate as being most similar to *quieto* 'calm, still' (3.97). It was followed by being related to *muerto* 'dead' (3.5), *espantado* 'scared' (3.03), *dormido* 'asleep' (3.03), *petrificado* 'petrified' (3.1) and *atónito* 'astounded' (2.97). *Mudo* 'mute' was rated as being most similar to *espantado* 'scared' (3.1), *quieto* 'calm, still' (2.9), *asombrado* 'astonished'(2.78), and *atónito* 'astounded' (2.6). However, used in *quedar(se)* + ADJ, *inmóvil* and *mudo* were mainly found in the data in contexts in which the subject became mute or motionless as a result of being surprised or scared. Because of this, they were proposed to relate to members of the *confuso / suspenso* clusters (Appendix 1, Sect. 3) and the *alegre / satisfecho* clusters (Chapter 5:, Sect. 5.3.1). The types belonging to the *confuso / suspenso* clusters will be discussed in more detail below.

Figure 37. Perceptual map of Questionnaire B; 6% smallest links.

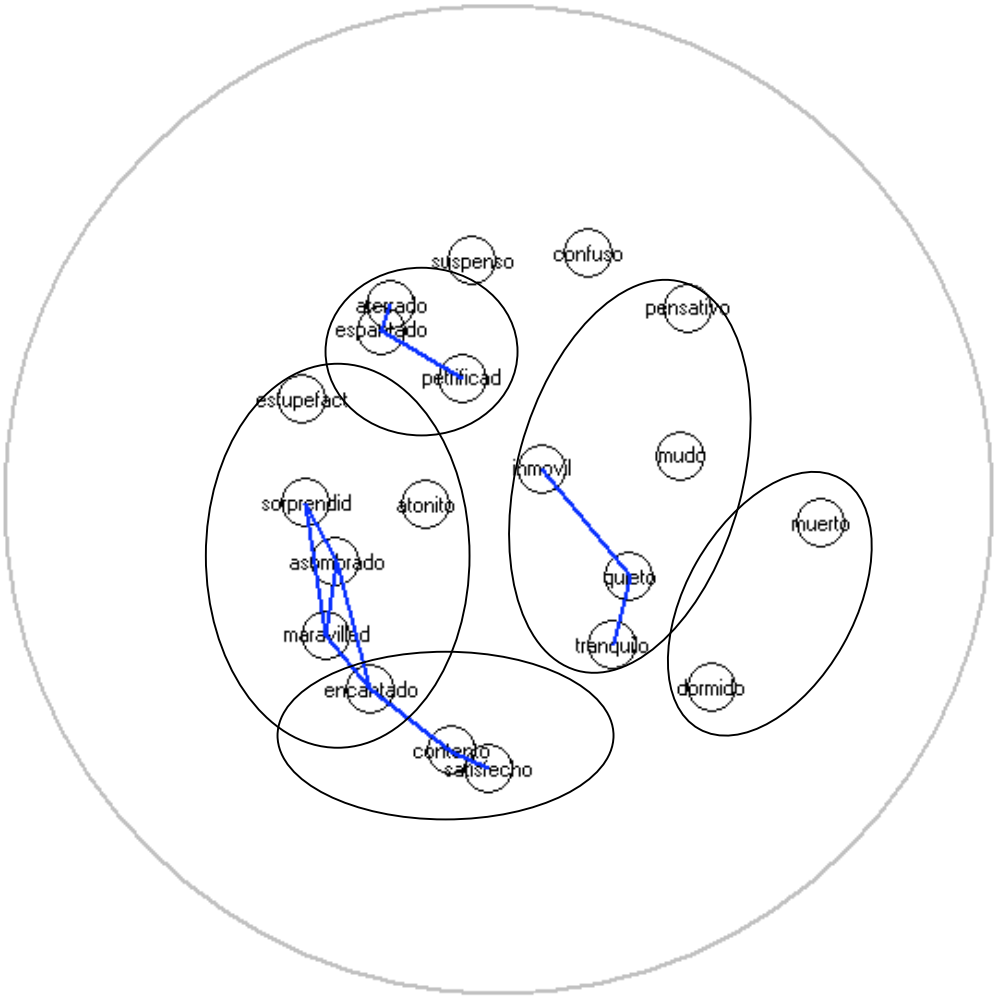


Figure 37 shows the Waern Links in solid lines between the 6% most similar pairs with the most similar being *asombrado* / *sorprendido* 'astonished / surprised' (4.8). The pattern of linked pairs mainly supports the exemplar clusters proposed in the previous chapter while presenting some unexpected results. As in Section 6.4.1, 'bubbles' have been added in order to delineate proposed categories of adjectives. No opposite values are proposed to exist between any of the pairs presented here and there is no proposed axis.

The types *aterrado* 'scared', *espantado* 'scared' and *petrificado* 'petrified' were encircled on the assumption that they were rated as highly similar because they all

convey a state of fear and form a potential category. The type *petrificado* is located closely to *inmóvil* 'motionless', another type that indicates a lack of motion. One of the unexpected outcomes of the perceptual map was that *suspenso* 'surprised' was placed closed to the 'scared' types while being away from other types conveying surprise. In other words, *aterrado*, *espantado*, and *petrificado* intervened in the potential category of types that denote surprise. This was partially due fact that *suspenso* was rated most closely to *aterrado* (3), *petrificado* (2.96), and *espantado* (2.9).

Intersecting with the 'fear' cluster is the cluster of adjective types that convey surprise. The three members in the center of this bubble, *sorprendido* 'surprised', *asombrado* 'astonished', and *maravillado* 'astonished' were all linked to one another by the Waern Links of the 6% highest similarity ratings. The related types, *atónito* 'astounded' and *estupefacto* 'stupefied' were located in close enough proximity to be encircled in order to represent the proposed category from Chapter 5:, but were not among the 6% most related pairs. *Encantado* 'enchanted' is proposed to be part of this cluster and part of the *contento / satisfecho* cluster in Figure 37. It has one 6% link to contento 'happy' and two 6% links to members of the 'surprise' cluster: *maravillado* and *asombrado.* This is an expected result because, in the data, *encantado* 'enchanted' appears in contexts that denote 'happiness' and in contexts that denote 'surprise'. In this chapter and in the analysis, *encantado* is included in clusters that reflect its usage: the *alegre / satisfecho* clusters (Chapter 5:, Sect. 5.3.1) and the *confuso / suspenso* clusters (Appendix 1, Sect. 3).

The 'bubble' encircling *muerto* 'dead' and *dormido* 'asleep' highlights the finding that they are in proximity to one another in perceptual space. This 'bubble' intersects

with the *inmóvil / quieto / tranquilo* ‘bubble’ in order to show similarities between members of these proposed clusters. Both *muerto* and *dormido* were rated as being similar to *inmóvil* ((3.5) and (3.03) respectively) and *dormido* was also rated as being similar to *tranquilo* ‘calm’ (3.03) and *quieto* ‘calm, still’ (3.2). However, the placement of *muerto* and *dormido* on the map is certainly influenced by the fact that the lowest similarity ratings found in the matrix involve pairings with these two types. The pairing *muerto / encantado* ‘dead / charmed’ had a rating of (1.1) as did the pairing *dormido / sorprendido* ‘asleep / surprised’. The fact that they are both outliers (i.e. generally dissimilar to other types) in the same area on the map indicates that, even though *muerto* and *dormido* only had a similarity rating of (2.16), they are still located in proximity to one another based on aggregate rating with other types. In other words, they are similar to many of the same types and dissimilar to many of the same types. This connection through sharing similarities and dissimilarities is supported by the fact that there are no intervening types between them.

6.5 Summary

As mentioned in Section 6.3.1, Borg & Groenen (1997: 3) propose four purposes for generating an MDS analysis. The two with the most relevance to the present study were (a) MDS as a method of analysis that presents similarity data as distances in a low-dimensional plot that can be visually inspected, and (b) a method of providing empirical evidence to support whether or not criteria proposed by researchers for distinguishing between items is mirrored in the perception of a larger group of subjects. The results of the analysis produced perceptual maps in which groupings of related adjectives coincided largely with the clusters of adjectives proposed in Chapter 5. This provides empirical

evidence that the criteria of similarity used in constructing these proposed clusters is mirrored in participant perception: members of proposed categories could be encircled in 'bubbles'.

The objective of submitting the pairings from Questionnaire A to an MDS analysis was to determine whether native speakers would rate the adjectives from each cluster as more similar to one another than to adjectives from different clusters. In general, the types from the proposed clusters were grouped together in a way that supported the proposed *solo* clusters, the *libre* clusters, and the *rico / pobre* clusters from the previous chapter. An analysis of the perceptual map, as shown in Figure 35, reveals that proposed opposites were located across the map from one another and that it was possible to apply an axis to the map between adjectives that portray positive and negative states.

The Waern Links provide further evidence of the cohesion of types proposed to be related. Among the links representing the 6% most highly similar pairs, eight out of ten connected members of the *solo* clusters to one another. The fact that *solo* was linked to more types than any other is evidence that it is the central category member and that it is semantically the most general. These 6% links also linked *pobre* 'poor' and *sin dinero*, as well as *rico* 'rich' and *heredero* 'heir'.

The objective of submitting the pairings of adjectives in Questionnaire B to an MDS analysis was to determine whether or not speakers rated the types in each proposed cluster as being similar to one another. Also, the goal of this questionnaire was to examine patterns of similarity regarding two types that are used in a variety of contexts: *mudo* 'mute' and *inmóvil* 'motionless'. A third goal, specific to the *alegre / satisfecho*

clusters, was to determine whether or not the analysis would support the proposed structure (both here and in Bybee & Eddington 2006) of these clusters in which unrelated types were linked by intermediary types.

With the exception of *suspenso* 'surprised', the perceptual map generated by the data from Questionnaire B placed types in a way that allowed members of proposed clusters to be encircled in a 'bubble'. In the *confuso / suspenso* clusters (Appendix 1, Sect. 3), I included adjectives denoting a state of fear (e.g. *espantado* 'scared') and adjectives denoting a state of surprise (e.g. *sorprendido* 'surprised'). This was based on the assertion that types in these clusters were similar because they both conveyed sudden reactions to something unexpected when used in *quedar(se)* + ADJ. Types denoting fear (e.g. *espantado* 'scared', *aterrado* 'scarred', and *petrificado* 'petrified') were grouped together in a way that allowed them to be encircled in a 'bubble' and were connected by 6% Waern Links.

Types denoting surprise were grouped in a way that allowed them to be grouped together as well (e.g. *estupefacto* 'stupefied', *sorprendido* 'surprised', *atónito* 'astounded', *asombrado* 'astonished', *maravillado* 'astonished', and *encantado* ' enchanted'). There were 6% links connecting the three types *sorprendido*, *asombrado*, and *maravillado*, and both *asombrado* and *maravillado* were connected to *encantado*. The latter type was also connected by a 6% to *contento* 'happy' which supports the proposition that it is a member of both categories. Even though the type *suspenso* was proposed to be a member of the 'surprised' cluster in the *confuso / suspenso* clusters, it was placed away from other types denoting surprise on the perceptual map.

As was proposed in Chapter 5, the types *mudo* 'mute' and *inmóvil* 'motionless' are similar to a variety of types. In the case of *inmóvil*, the central position on the perceptual map is evidence of this. The many similarities that these adjectives have to other types supports the proposition that they should be members of both the *confuso / suspenso* clusters and the *alegre / satisfecho* clusters.

This MDS analysis also supports the notion that there are links between clusters in which unrelated members (e.g. *contento* 'happy', and *muerto* 'dead') could be related by intervening types. The key intervening types in this chain are *tranquilo* 'calm' and *quieto* 'still, calm' which allow for types such as *dormido* 'asleep', *inmóvil*, and *mudo* to play a role in the linkage that connects *contento* 'happy' and *satisfecho* 'happy' to *muerto* 'dead'.

There is one final point regarding types that link otherwise unrelated types. They show how it is possible that, through some intervening link, all categories may be related. The central members provide cohesion by attracting similar types while the intervening ones could be pathways for expansion.

Chapter 7: Conclusion

The preceding chapters have outlined the results and findings concerning emergent adjective usage in the change-of-state construction *quedar(se)* + ADJ over time. The exemplar model was used in order to account for diachronic tendencies of adjective categories including emergence, expansion, and contraction. The database spans seven centuries, from the 1200's to the 1800's, and some commentary was made regarding the 1900's using data from Bybee & Eddington (2006). Data from the *Corpus del español* (Davies 2002-) was used in order to supplement commentary on the 1900's and was used to inform the analysis as well. This is the first study to implement the exemplar model to account for the emergence of semantic categories over a long span of time. In general terms, the results are two-fold: findings that reveal diachronic patterns of distribution of adjectives used in *quedar(se)* + ADJ, and findings that contribute to the general understanding of the role of constructions in language change.

7.1 Overview of quedar(se) *+ ADJ*

An analysis of the data has shown that *quedar(se)* + ADJ underwent a rise in standardized overall frequency in my data, as measured by occurrences per 10,000 words, before undergoing a loss in overall frequency. It went from occurring in the data at a rate of 0.42 tokens per 10k words in the 1200's, to reaching a peak of overall frequency in the 1600's of 6 tokens per 10k words, and falling to 2.64 occurrences per 10k words in the 1800's. Using Bybee & Eddington's data, I calculated that the overall frequency had fallen even farther in the 1900's to 1.83 tokens per 10k words (see Sect. 4.2.1). This pattern of standardized overall frequency through the span of the data analyzed resembles

a ‘bell curve’ and is taken as evidence that trends of usage from one century affect the trends of usage in subsequent centuries; it is not a random pattern.

These fluctuations in overall frequency may have been affected by frequency shifts in other verb + adjective expressions of ‘becoming’. The change-of-state expression *fincar(se)* + ADJ was much more frequent than *quedar(se)* + ADJ in the 1200’s. By the 1400’s, *quedar(se)* + ADJ had surpassed the former construction in overall frequency as it was in a process of gaining frequency while *fincar(se)* + ADJ was undergoing a loss in frequency. Many adjectives found to be used in *fincar(se)* + ADJ went on to be used in *quedar(se)* + ADJ, while others showed semantic similarity to adjectives comprising the proposed categories of this latter construction.

In a similar interaction, as *quedar(se)* + ADJ began to undergo a loss in overall frequency, many adjective types went on to be used in other expressions of ‘becoming’. Many adjectives that occurred in *quedar(se)* + ADJ were demonstrated to co-occur with *poner(se)* + ADJ from the 1600’s to the 1800’s and there is evidence that this latter expression of ‘becoming’ was seeing a gain in overall frequency (4.2.1.3).

Both of these scenarios show that there are shifts in both the overall frequency of expressions of ‘becoming’ and in the distribution of adjectives among these constructions. Adjectives from one change-of-state construction come to be used with another change-of-state construction over time. There is a correlation between token and type frequency; as the construction *quedar(se)* + ADJ shows evidence of extension by including new types in usage, it also shows evidence of a simultaneous increase in token frequency.

7.2 Formulaic language and frequency effects

The role of frequently occurring combinations in the development of adjective categories in *quedar(se)* + ADJ has been highlighted in several ways. In order to study frequent sequences, two different measures of formulaicity were proposed, a conventionalized instances of constructions (CIC) threshold and a prefab threshold (See Chapter 4:, Section 4.2.1.1). The CIC threshold represents types that have twice the average token frequency and is useful in providing a general figure that is readily applicable to a wide group of types. On the other hand, the prefab threshold includes only combinations with no intervening matter, and compares the standardized frequency of the combination to the standardized frequency of both the verb and the adjective. It was only applied to two combinations (*quedar(se) solo* 'to be left alone' and *quedar(se) libre* 'to be set free') and was instrumental in advancing the claim that prefabs have longevity. While the CIC threshold includes prefabs, the more specific prefab threshold would exclude some CIC's. Both thresholds of formulaic combinations provide analysis tools that help support the claim that frequently occurring combinations are the central members of categories and that there is a correlation between the token frequency of the central member and the type frequency of the category to which it belongs. In fact, the correlation between type and token frequency could be an effect of gradient category structure. As a category becomes more productive, it initially attracts types related to the central member before extending to other types. In a waning phase, the central member endures even as other types disappear in the data.

The exemplar model, as discussed in Section 2.1, is an instance-based model of categorization in which categories are formed through a speaker's accumulated

experience with language; frequently experienced items gain strength in representation and serve as the central members of gradient categories. This study, it follows, has provided evidence that frequently occurring combinations of *quedar(se)* + ADJ are the central members of gradient categories, and are surrounded by related adjective types in a hypothetical cognitive space. These categories, like many of their central members, exhibit longevity.

In investigating the ways in which previous linguistic usage affects subsequent usage, the study of prefabs has shown that these may retain prefab status over many centuries. As mentioned in Chapter 5:, Section 5.2.1.8, for example, the prefab *quedar(se) solo* 'to be left alone' appeared in the data above the CIC threshold in the entire span of the data studied, and into the 1900's, showing a steady increase in frequency. It appeared above the prefab threshold from the 1500's to the 1900's. In other words, prefabs, CIC's, and formulaic language in general have longevity.

Even in the cases where prefabs emerged, or faded, in the span of the data, they did so in a way that related to previous usage. In the case of *quedar(se) libre* 'to be set free' (Sect. 5.2.2.8), it showed a trend of gaining in standardized frequency from the 1200's to the 1500's before showing a reduction of frequency in my data. There was a similar pattern in the *Corpus del español* (Davies 2002-) in which the peak occurred in the 1600's; also, it was a prefab in those data from the 1300's to the 1800's. The overarching pattern resembled a bell curve indicating that it is either in a trend of emerging or waning in usage and not random; trends of usage from one century affect trends in subsequent ones.

Prefabs do not exist in a vacuum and have an impact on the categories to which they belong. As the central members gain in token frequency, there is a tendency for the pertinent category to extend by undergoing a gain in type frequency. For example, the prefab *quedar(se) solo* steadily increases in token frequency[30]. As it reaches its peak in the 1800's (0.3 occurrences per 10k words), its category also reaches a peak in type frequency in which types from the *solo* clusters account for 11.5% of all types in the data (see 5.2.1.8). One of the reasons that categories may be extended as the central member gains in token frequency is that, by occurring more frequently in usage, a type simply has more opportunities for being used as an analogical model of extension.

Also, this study informs the mechanisms of change regarding constructions in which the context of usage plays a central role. In a way that is similar to grammaticization, as proposed by Bybee, Perkins, & Pagliuca (1994), the context of usage is the nexus of change. As a verb + adjective combination is used in one context, it can be extended through metaphor or inference to other contexts. Said differently, combinations extend through analogical extension. Frequently occurring combinations, as stated in the previous paragraph, have more opportunities to participate in this process as models of extension. The fact that frequent combinations may generate more new usages is also evidence that they should be the central members of exemplar categories.

7.3 Categories over time

One of the central findings of this thesis is that it is possible to organize the adjectives found in the data for each century into plausible categories based on their semantic similarity. This supports the proposal that these categories exist, and are

[30] This was despite an idiosyncratic drop in the 1700's that was also confirmed in Davies' (2002-) data.

coherent, based on this observed similarity. Despite the fact that they change over time as new members emerge and old ones are discontinued in usage, these categories tend to span several centuries.

The vehicle for expansion is analogical extension in which novel items emerge in usage through analogy to experienced types that serve as a model in the production of the novel item. While I have proposed that frequently occurring members have a higher probability of serving as a model for extension because they are experienced more often, infrequent members could also serve as analogical models. In fact, one of the advantages to using the exemplar model is that it allows for category expansion to occur through analogy to marginal or redundant features (Bybee & Eddington 2006, Chandler 2002). Consequently, one of the ways in which many categories expanded was through the generalized forms *quedar(se) sin* + NOUN and *quedar(se) con* + NOUN which supplied a sort of 'template' which allowed for the emergence of many new types. This allowed for expansion based on form, by allowing nouns to be used in the construction as the objects of prepositional phrases, and meaning.

There are also internal changes in the categories in which the central member shifts from one type to another. This is seen when the central member of a category is overtaken in token frequency by another member of the category. In the *alegre / satisfecho* clusters (Sect. 5.3.1), the central member goes from *alegre* 'happy', to *contento* 'happy' in the 1500's, and to *satisfecho* 'satisfied' from the 1600's onward.

7.4 Further considerations

This investigation has provided an in-depth analysis of the adjectives found in the change-of-state construction *quedar(se)* + ADJ in seven centuries. Part of the purpose of

the research was to develop a method for analyzing potential categories and investigating patterns of emergence. While many trends regarding prefabs, frequency, and category structure were uncovered, future research could look more closely at some of the specific issues in a diachronic setting.

One of the findings was that *quedar(se)* + ADJ inherited adjectives from the waning construction *fincar(se)* + ADJ in the earlier. It would be revealing to look at early expression of 'becoming' from the 1200's to the 1400's in order to explore this in more detail. This analysis could include other archaic verbs of 'becoming' such as *trocar(se)*. Looking a little bit farther back, it would be interesting to see if there were related constructions in Vulgar Latin. Furthermore, it would be informative to look at expressions of 'becoming' in related languages. In Portuguese, *fincar(se)* is used in a similar manner to *quedar(se)* in Spanish and is used to express a change of state[31].

On the more modern end of the data span, it was found that as *quedar(se)* + ADJ began to drop in overall frequency from the 1600's onward, many of the adjectives used in this construction were co-occurring with the change-of-state construction *poner(se)* + ADJ. This is indicative of an emergent process in which adjectives previously associated with one expression of 'becoming' could come to be associated with a different one. It would be insightful to study the shifts, or co-occurrence, of adjectives among other expressions of 'becoming' to see how this correlated with the overall frequency of these expressions. For example, it would be worthwhile to investigate the hypothesis that the rise in frequency of *poner(se)* + ADJ is related to the loss in frequency of *quedar(se)* + ADJ. This study has, after all, shown the way in which this type of analysis reveals diachronic trends in formulaic language. It would be illuminating to apply it to related

[31] Agripino Silveira, p.c.

change-of-state constructions in order to study their interaction over time and better understand language change and endurance.

One final consideration must be mentioned. As was observed in the analysis of the data, the construction *quedar(se)* + ADJ is in a trend of losing frequency. If it follows a path similar to that of its predecessor, *fincar(se)* + ADJ, it will eventually dissapear from usage, even though this may not happen for a few hundred years. This raises the question as to whether or not there is another expression of 'becoming' that may emerge. If this were the case, it would be informative to look at other verbs that express 'remaining' to see if they are being used with adjectives in ambiguous contexts that could possibly be interpreted as a change of state. As was mentioned in Section 4.2.1.2, *permanecer* 'to stay, remain' could be a possible candidate for being used as a 'become' verb. If it is being used with adjectives similar to the ones used with *quedar(se)* + ADJ, and some of these usages express 'becoming', it is possible that a new change-of-state construction (i.e. *permanecer(se)*) is in the early stages of gaining traction and overcoming *quedar(se)* + ADJ.

References

Baayen, R. Harald. 1993. On frequency, transparency and productivity. In *Yearbook of Morphology 1992*, Geert Booij and Jaap van Marie (eds.), 181-208. Kluwer Academic.

Balasch, Sonia. 2008. Construction of *poner(se)* + ADJ: 17th – 19th century. Final project for graduate seminar *Frequency of Use and Emergent Grammar*, Professor Bybee, UNM, Spring 2008.

Bod, Rens. 2006. Exemplar-based syntax: How to get productivity from examples. *The Linguistic Review* 23(3).275-290.

Borg, Ingwer & Patrick Groenen. 1997. *Modern Multidimensional scaling: Theory and applications.* New York: Springer.

Butt, John, & Benjamin, Carmen. 2004. *A new reference grammar of modern Spanish*. 4th ed. New York : McGraw-Hill

Bybee, Joan. 1985. *Morphology: A study of the relation between meaning and form.* Philadelphia: John Benjamins.

Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, Joan. 2002a. Phonological Evidence for Exemplar Storage of Multiword Sequences. *SSLA*. 24.215-221.

Bybee, Joan. 2002b. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*. 14(3).261-290.

Bybee, Joan. 2003. Mechanisms of change in grammaticization: the role of frequency. In *Handbook of Historical Linguistics,* Brian Joseph and Richard Janda (eds.), 602-623. Oxford: Blackwell Publishers.

Bybee, Joan. 2006. From usage to grammar: the mind's response to repetition. *Language* 82(4).529-551.

Bybee, Joan L., & David Eddington. 2006. A Usage-based Approach to Spanish Verbs of 'Becoming'. *Language*, 82(2).323-355.

Bybee, Joan, Revere Perkins, & William Pagliuca. 1994. *The Evolution of Grammar; Tense, Aspect, and Modality in the Languages of the World.* Chicago and London: University of Chicago Press.

Bybee, Joan and Sandra Thompson. 2000. Three frequency effects in syntax. *Berkeley Linguistic Society* 23.65-75.

Bybee, Joan and Rena Torres Cacoullos. 2009. The role of prefabs in grammaticization: How the particular and the general interact in language change. In *Formulaic language Vol. 1: Structure, Distribution, and Historical Change.* R. Corrigan, H. Ouali, E. Moravcsik, & K. M. Wheatley (eds.) 187-218. Amsterdam / Philadelphia: Benjamins..

Chandler, Steve. 2002. Skousen's analogical approach as an exemplar-based model of categorization. In *Analogical Modeling: An exemplar-based approach to language,* R. Skousen, D. Lonsdale, & D. B. Parkinson (eds.), 51-105. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Chomsky, A. N. 1957. *Syntactic Structures*. The Hague: Mouton. Reprint. Berlin and New York (1985).
Coste, Jean, & Augustín Redondo. 1965. *Syntaxe de l'espagnol moderne*. Paris: Société d'Edition d'Enseignement Superieur.
Crespo, Luis A. 1949. To become. *Hispania* 32.210-12
Croft, William. 2001. *Radical Construction Grammar: syntactic theory in typological perspective.* Oxford: Oxford University Press.
Croft, William. 2003. *Typology and Universals, 2nd ed.* (Cambridge Textbooks in Linguistics.) Cambridge: Cambridge University Press.
Croft, William. 2007. Exemplar Semantics. Draft, to appear. Accessed from website, http://www.unm.edu/~wcroft/ in May 2008. Originally presented at Paper presented at 8th Conference on Conceptual Structure, Discourse, and Language : Language in Action (CSDL8) at UCSD. November 3-5 2006, La Jolla, CA.
Croft, William, & Keith T. Poole. 2008. Inferring universals from grammatical variation: multidimensional scaling for typological analysis. *Theoretical Linguistics* 34.1-37.
Davies, Mark. 2002-present. Corpus del Español (100 million words, 1200s-1900s). Available online at www.corpusdelespanol.org.
Diccionario de la Real Academia Española (DRAE), 23rd edition. 2005. www.rae.es. Accessed Fall 2006 to Spring 2009.
Eberenz, Rolf. 1985 Aproximación structural a los verbos de cambio en Iberorromance. *Linguistique comparée et typologie des langues romanes*, ed. By Jean-Claude Bouvier, 460-75. Aix en Provence: Université de Provence.
Eddington, David. 1999. On 'becoming' in Spanish: a corpus analysis of verbs expressing change of state. *Southwest Journal of Linguistics* 18.23-46.
Eddington, David. 2002. Disambiguating Spanish change of state verbs. *Hispania* 85.921-29
Erman, Britt, & Beatrice Warren. 2000. The idiom principle and the open choice principle. *Text* 20(1).29-62.
Faulhaber, Charles B., Angel G. Moreno, Antonio C. Ocaña, & Oscar P. Rodríguez. 2002. Bibliografía Española de Textos Antiguos. Volume 2002, Number 2 (July). http://sunsite.berkeley.edu/Philobiblon/phhmbe.html. Accessed in Spring 2007.
Faulhaber et. al. 1984. *Bibliography of Old Spanish Texts*, 3rd Ed. Madison; Hispanic Seminary of Medieval Studies.
Fente, R. 1970. Sobre los verbos de cambio o 'devenir'. *Filología Moderna* 38.157-72.
Fillmore, Charles J. 1977. Topics in lexical semantics. In *Syntax and Semantics 8: Grammatical Relations*, R. Cole (ed.), 59-81. New York: Academic Press.
Gahl, Susanne, & Alan C. L. Yu. 2006. Introduction to the special issue on exemplar-based linguistics. *The Linguistic Review*, 23(3).213-216.
Givón, T. 1979. *On understanding grammar*. New York: Academic Press.
Goldberg, Adele. 1995. *Constructions: A construction grammar approach to argument structure.* Chicago: University of Chicago Press.
Goldberg, Adele. 2003. Constructions: a new theoretical approach to language. *Trends in Cognitive Sciences* 7(5).219-24.
Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. New York: Oxford University Press.

Greaves, Chris. 2005. Concapp Windows Application V4. http://www.edict.com.hk/PUB/concapp/. Downloaded in Spring 2007.
Hay, Jennifer. 2001. Lexical frequency in morphology: Is everything relative? *Linguistics* 39.1041-1070.
Heady, Ronald. 2007. PERMAP, PERceptual MAPping Software, 11.6. http://www.ucs.louisiana.edu/~rbh8900.
Heady, Ronald B., & Jennifer L. Lucas. 2007. PERMAP 11.6 Operation Manual. http://www.ucs.louisiana.edu/~rbh8900.
Hopper, Paul J. 1987. Emergent grammar. *Proceedings of the Berkeley Linguistics Society* 13.139-157.
Hopper, Paul J. and Sandra A. Thompson. 1984. The discourse basis for lexical categories in universal grammar. *Language* 60(4).703-752.
Hintzman, Douglas L. 1986. 'Schema abstraction' in a multiple-trace memory model. *Psychological Review* 4.528-551.
Israel, Michael. 1996. The way constructions grow. *Conceptual structure, discourse and language*, Adele Goldberg (ed.), 217-31. Stanford, CA: SSLI Publications.
Kellog, Michael. 1999. *WordReference.com.* Vienna, Virginia, USA. <www.wordreference.com> accessed Spring 2008.
Kemmer, Suzanne. 1993. The middle voice: Typological studies in language. Amsterdam / Philadelphia: John Benjamins.
Kemmer, Suzanne. 2003. Human cognition and the elaboration of events: Some universal conceptual categories. *The new psychology of language: Cognitive and functional approaches to language structure*, Michael Tomasello (ed.), 89-118. Mahwah, NJ: Lawrence Erlbaum.
Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago & London: University of Chicago Press.
Langacker, Ronald. 1987. *Foundations of cognitive grammar, vol. 1: Theoretical prerequisites.* Stanford, CA: Stanford University Press.
Lapesa, Rafael. 1980. *Historia de la lengua española.* Madrid: Editorial Gredos.
Lathrop, Tom. 1996. *The Evolution of Spanish.* Newark, Delaware: Juan de la Cuesta.
Lipski, John. 1984. On the Weakening of /s/ in Latin American Spanish. *Zeitschrift Für Dialektologie und Linguistik* 51(1).31-43.
Logan, Gordon D. 1988. Toward an instance theory of automatization. *Psychological Review* 95.492-527.
Medin, Douglas L., & Marguerite M. Schaffer. 1978. Context theory of classification learning. *Psychological Review*. 85(3).207-38
Murphy, Gregory L. 2002. *The Big Book of Concepts.* Cambridge, MA: MIT Press.
Nofosky, Robert M. 1986. Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General,* 115.39-57.
Nofosky, Robert M. 1988. Similarity, Frequency, and Category Representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 14(1).54-65.
Nofosky, Robert M. 1992. Similarity scaling and cognitive process models. *Annual Review of Psychology,* 43.25-53.
Nofosky, Robert M., & Thomas J. Palmeri. 1997. An exemplar-based random walk model of speeded classification. *Psychological Review,* 104.266-300.

Noonan, Michael. 1998. Non-structuralist syntax. In *Functionalism and formalism in linguistics: General papers*, Vol. 1, Michael Darnell, Edith Moravcsik, Frederick Newmeyer, Michael Noonan and Kathleen Wheatley (eds.), 11-31. Amsterdam / Philadelphia: John Benjamins.

O'Neill, John. 1999. *Electronic Texts and concordances of the Madison Corpus of Early Spanish Manuscripts and Printings.* Madison and New York: Hispanic seminary of Medieval Studies.

Penny, Ralph. 2000. *Variation and Change in Spanish*. Cambridge: Cambridge University Press.

Pierrehumbert, J. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In *Frequency and the emergence of linguistic structure*, J. Bybee & P. Hopper (eds.), 137-157. Amsterdam: Benjamins.

Pountain, Christopher J. 2001. *A History of the Spanish Language through Texts.* London: Routelage.

Real Academia Española, RAE. 2005. Diccionario de la lengua española. 23rd Ed. <www.rae.es> accessed Spring 2008.

Rosch, Eleanor. 1978. Principles of categorization. In *Cognition and categorization*, Eleanor Rosch and Barbara L. Lloyd (eds), 27-48. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schwegler, Armin, & Juergen Kempff. 2007. *Fonética y fonología españolas, 3rd ed.* Hoboken, NJ: John Wiley & Sons, Inc.

Skousen, Royal. 2002. An overview of Analogical Modeling. In *Analogical Modeling: An exemplar-based approach to language*, R. Skousen, D. Lonsdale, & D B. Parkinson (eds.), 11-26. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Skupin, André, & Barbara Buttenfield. 1997. Spatial Metaphors for Display of Information Spaces. Proceedings, AUTO-CARTO 13, 116-125.

Smith, E. E. & Osheron D. N. 1984. Conceptual combination with prototype concepts. *Cognitive Science* 8.337-361.

Smith, K. Aaron. 2001. The role of frequency in the specialization of the English anterior. In *Frequency and the emergence of linguistic structure,* Joan Bybee and Paul J. Hopper (eds), 362-382. Amsterdam / Philadelphia: John Benjamins.

Stalans, Loretta J. 1995. Multidimensional scaling. In *Reading and understanding multivariate statistics*, L. G. Grimm & P. R. Yarnold (Eds.), 137-168. Washington, DC: American Psychological Association.

Taylor, John R. 1995. *Linguistic Categorization*. Oxford: Oxford University Press.

Tomasello, Michael (ed). 1998. *The new psychology of language: Cognitive and functional approaches to language structure*, vol. 1, 2 vols. Mahwah, NJ: Lawrence Erlbaum Associates.

Tomasello, Michael (ed). 2003. *The new psychology of language: Cognitive and functional approaches to language structure*, vol. 2, 2 vols. Mahwah, NJ: Lawrence Erlbaum Associates.

Torres Cacoullos, Rena. 2000. *Grammaticization, synchronic variation, and language contact. A study of Spanish progressive –ndo constructions.* Amsterdam: John Benjamins.

Tottie, Gunnel. 1991. Lexical diffusion in syntactic change: Frequency as a determinant of linguistic conservatism in the development of negation in English. In *Historical English syntax*, Dieter Kastovsky (ed), 439-467. Berlin: Mouton de Gruyter.
Travis, Catherine E. and Agripino S. Silveira. 2009. The role of frequency in first-person plural variation in Brazilian Portuguese: *Nós* vs. *a gente*. Studies in Hispanic and Lusophone Linguistics 2(2)
Wedel, Andrew B. 2006. Exemplar models, evolution and language change. *The Linguistic Review*, 23(3).247-274.
Wilson, Damián. 2006. From 'Remaining' to 'Becoming' in Spanish: A diachronic usage-based approach. Paper presented at High Desert Linguistics Society (HDLS) 7, Albuquerque. November 9-11, 2006.
Wilson, Damián. 2007. From ‘remaining’ to ‘becoming’ in Spanish: the analogical extension of *quedar(se)* + ADJ. Paper presented at UWM Linguistics Symposium on Formulaic Language, Milwaukee, MN. April 18-21, 2007.
Wilson, Damián. 2009. From ‘remaining’ to ‘becoming’ in Spanish: The role of prefabs in the development of the construction *quedar(se)* + ADJECTIVE. In *Formulaic language Vol. 1: Structure, Distribution, and Historical Change.* R. Corrigan, H. Ouali, E. Moravcsik, & K. M. Wheatley (eds.), 273-295. Amsterdam / Philadelphia: Benjamins.
Wray, Alison, & Michael R. Perkins. 2000. The functions of formulaic language: an integrated model. *Language & Communication* 20.1-20.
Zuraw, Kie. 2003. Probability in Language Change. In *Probabilistic Linguistics*, Bod, Rens, Jennifer Hay, & Stefanie Jannedy (eds.), 139-176. Cambridge: MIT Press.

List of appendices

Appendix 1: Adjective clusters and tables not analyzed in Chapter 5:

1. Changes in social status

Table 29, 1200's: Adjective types denoting social states

Conflict:	
desafiado 'challenged'	1
Social Status:	
por señor 'as boss'	2
por sucesor 'as sucessor'	1
por adelantado 'as military boss'	1
por caudillo 'as leader'	1
por alevoso 'as a traitor'	

Table 30, 1300's: Adjectives reflecting social states

Conflict:	
vencido 'defeated'	2
Adjectival nouns:	
rey 'king'	1
señor 'boss'	1
Social Status:	
por señor 'as boss'	5
por rey 'as king'	3
por sus vasallos 'as their vassals'	1

Table 31, 1400's: Adjectives reflecting social states

Conflict:	
desafiado 'challenged'	3
vencedor 'victor'	3
concorde 'in agreement'	5
concordado 'in agreement'	1
concertado 'in agreement'	1
en gran diferencia 'at odds'	1
enmendador 'rectifying'	1
en paz 'in peace'	1
en amistad 'in friendship'	1
Adjectival nouns:	
rey 'king'	1
alcalde 'mayor'	1
doncella 'damsel'	1
señor 'boss'	1
maestro 'teacher'	1
condesa 'countess'	1
sucesor 'succesor	4
Social Status:	
por señor 'as boss'	4
por rey 'as king'	4
por su reina 'as their queen'	1
por regidores 'as leader'	2
por gobernador 'as governor'	3
por sus vasallos 'as their vassals'	1
por sus tutores 'as their tutors'	2
por enxiemplo 'as (an) example'	1
en sucesión 'in succession'	1

Table 32, 1500's: Adjectives reflecting social states

Conflict:	
vencido 'defeated'	1 (other 2 = 'tired')
deshonrado 'dishonored'	3
por vencido 'as defeated'	1
escarmentado 'to have learned lesson'	1
vencedor 'victor'	2
victorioso 'victorious'	1
vengado 'avenged'	1
por vencedor 'as victor'	2
concertados 'in agreement'	1
apareado 'equalized'	1
en buena paz 'in good peace'	2
en amistad 'in friendship'	1
infamado 'defamed'	1
enmendado 'rectified'	1
Adjectival nouns:	
amigo 'friend'	1
dueño 'owner'	1
monja 'nun'	1
señor 'boss'	1
noble 'noble'	1
Social Status:	
por señora 'as boss'	1
convertido en nuevo ser 'converted into a new being'	1
acreditado 'accredited'	1
limpio de honor 'clean of honor'	2
consagrado 'consecrated'	2
reverenciado 'revered'	1
Social Obligation:	
obediente 'obedient'	2
servido 'served'	1
muy en su gracia 'very much in his grace'	1

Table 33, 1600's: Adjectives reflecting social states

Conflict:	
vencido 'defeated'	4
rendido 'defeated'	4
descubiertos de amparo 'uncovered from protection'	1
triunfante 'triumphant'	1
honrado 'honored'	1
vengado 'avenged'	4
endiosado 'elevated to divinity'	
Adjectival nouns:	
amantes 'lovers'	1
hermano 'brother'	1
Social Status:	
por rey 'as king'	1
por señor 'as boss'	1
por tu esposo 'as your husband'	1
instituído por visorrey 'instituted as viceroy'	1
con nombre de valerosas 'w/ a valiant, honorable name'	1
Social Obligation:	
ocupado 'occupied'	1
servido 'served'	1

Table 34, 1700's: Adjectives reflecting social states

Conflict:	
vencido 'defeated'	1
escarmentado 'have learned a lesson'	2
oprimido 'oppressed'	3
sin libertad 'without liberty'	1
sin apoyo 'without support'	1
vengado 'avenged'	1
victorioso 'victorious'	1
glorioso 'glorious'	1
lisonjeado 'admired'	1
Social obligation:	
acreditado 'reputable'	1
con este cargo 'with this reponsibility'	1
en cargo de satisfaceros 'in charge of satisfying you all'	1
de acuerdo 'in agreement'	2
Reputation:	
despreciado 'despised'	1
aborrecido 'loathed'	1
lisonjeado 'admired'	1
tildado 'denigrated'	
Other:	
descubierto 'discovered'	1
conocido 'known'	1
despedido 'taken leave of, fired'	2
servido de cierto papel 'to have received a certain paper'	1
dispuesto 'willing'	1
deseoso 'eager'	1
aprovechado 'benefited'	1
colocado 'located'	1
Adjectival nouns:	
presbítero 'ordained clergy'	1
predicador sabatino 'sabbatine preacher'	1
por dueño 'as owner'	1

Table 35, 1800's: Adjectives reflecting social states

Conflict:	
deshonrado 'dishonored'	1
vencido 'defeated'	2
escarmentado 'have learned a lesson'	1
triunfante 'triumphant'	1
repletos de honor 'replete with honor'	1
airoso 'honorable'	
Adjectival nouns:	
amigo 'friend'	1

2. The *triste* clusters

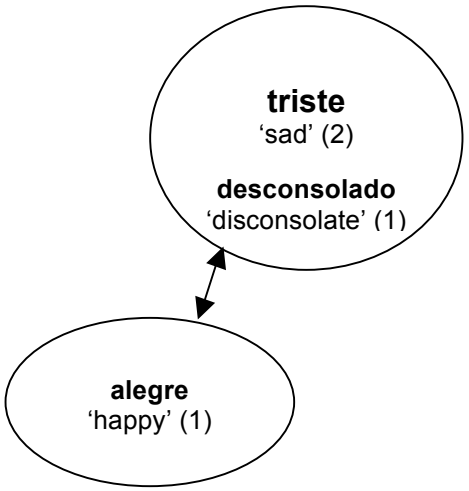


Figure 38, 1200's: the *triste* clusters

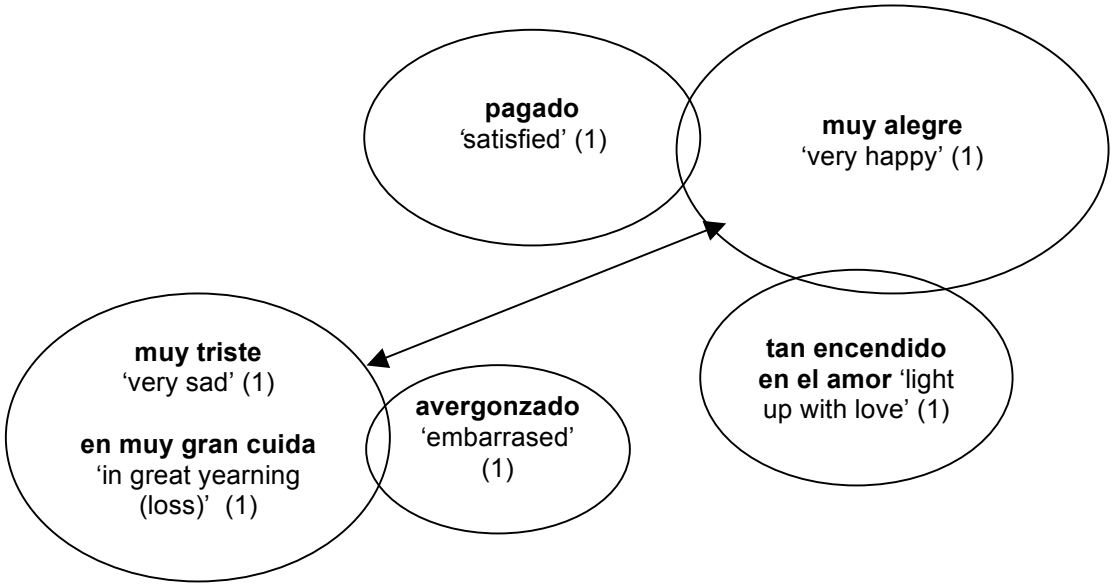


Figure 39, 1300's: the *triste* clusters

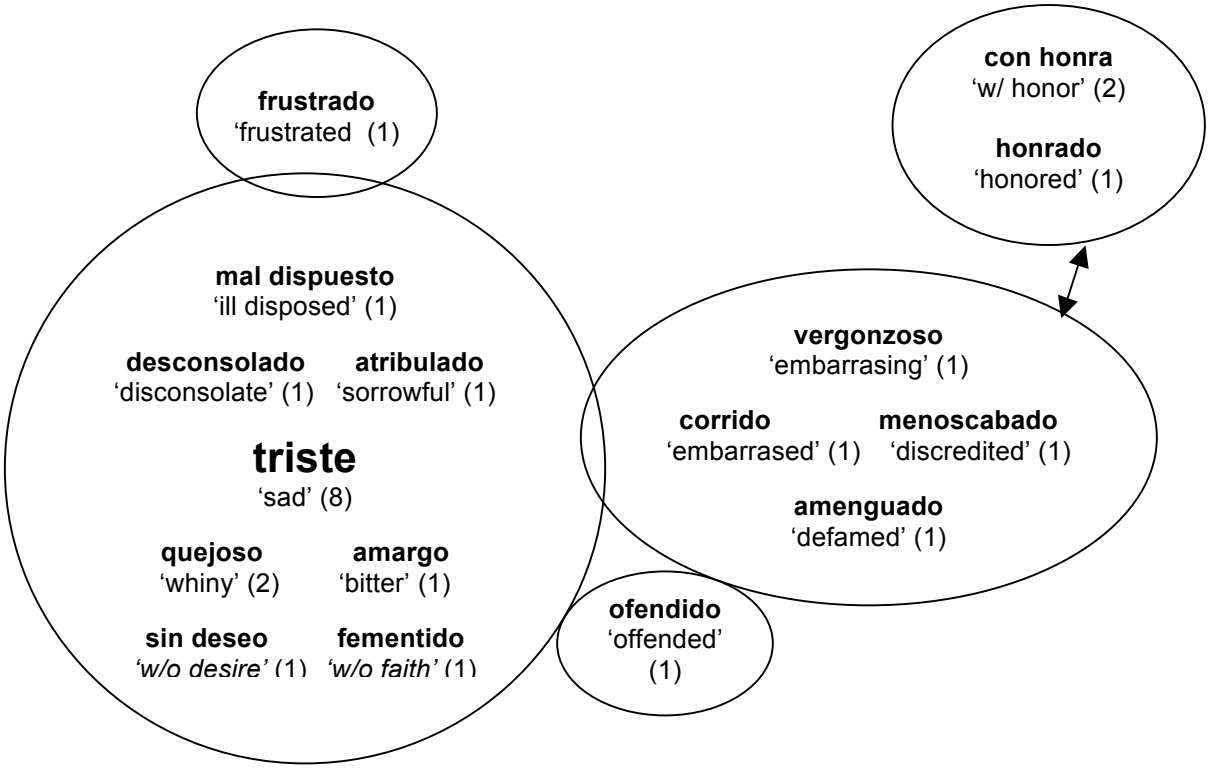


Figure 40, 1400's: the *triste* clusters

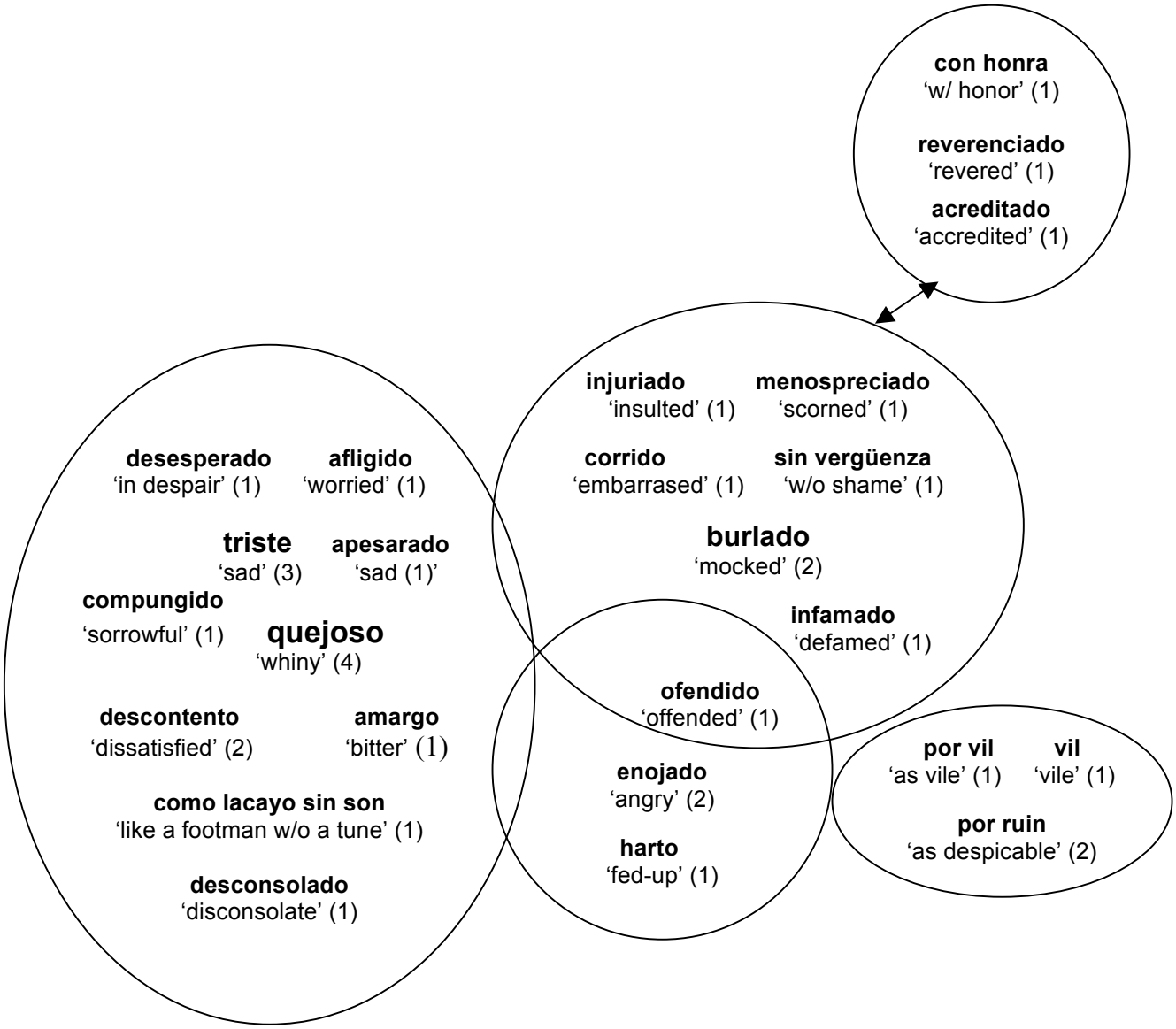


Figure 41, 1500's: the *triste* clusters

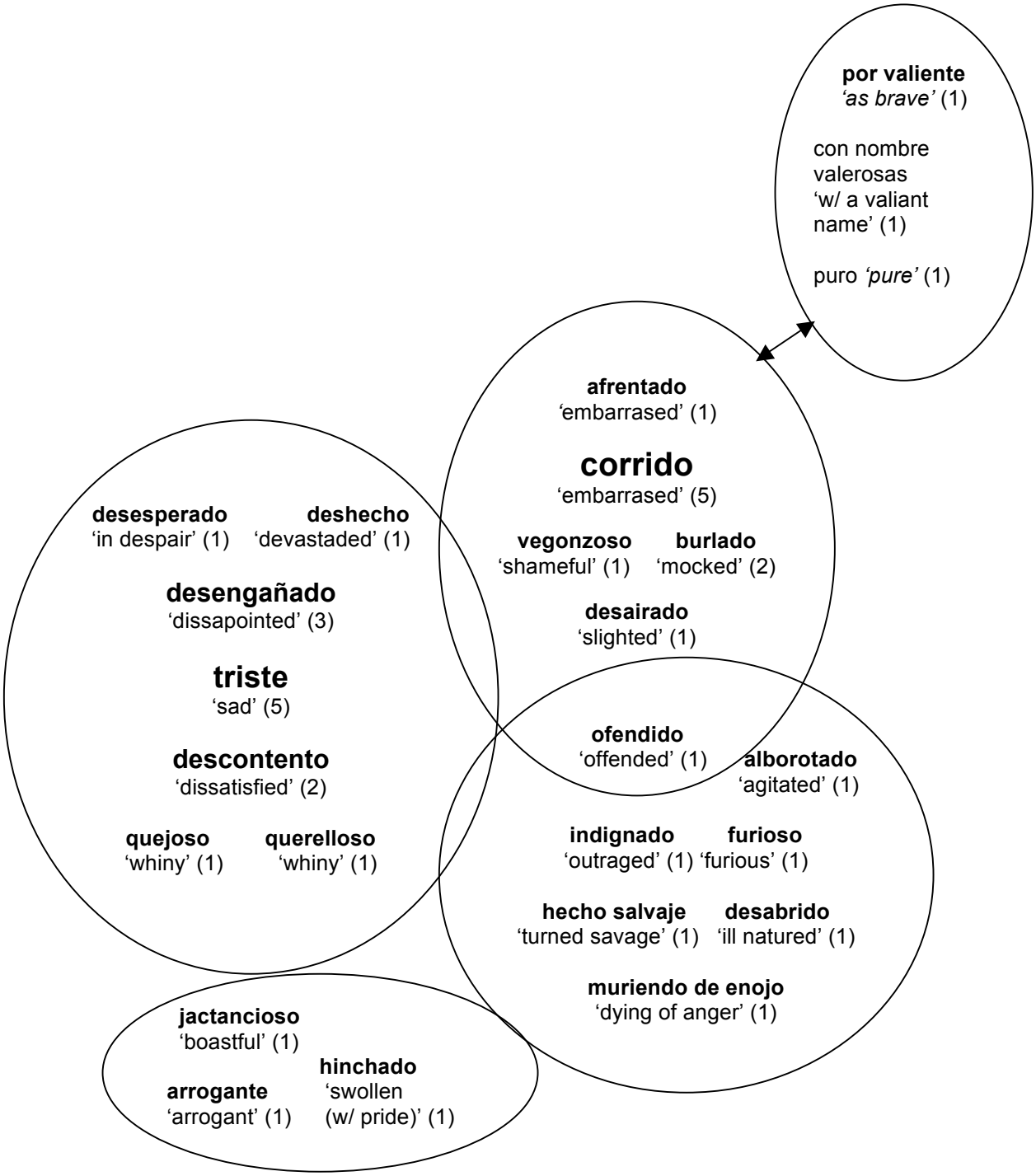


Figure 42, 1600's: the *triste* clusters

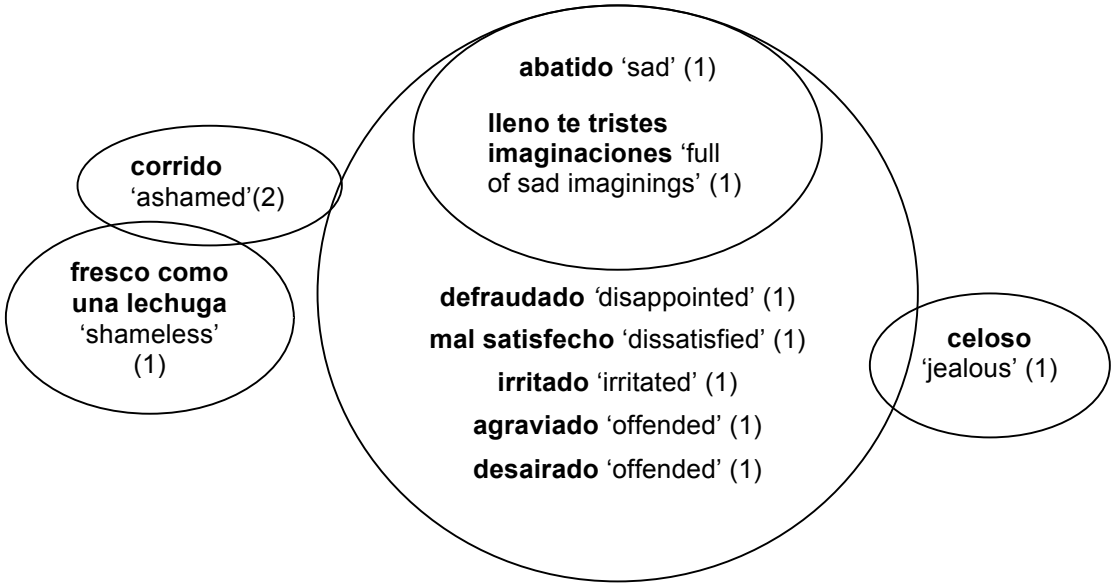


Figure 43, 1700’s: the *triste* clusters

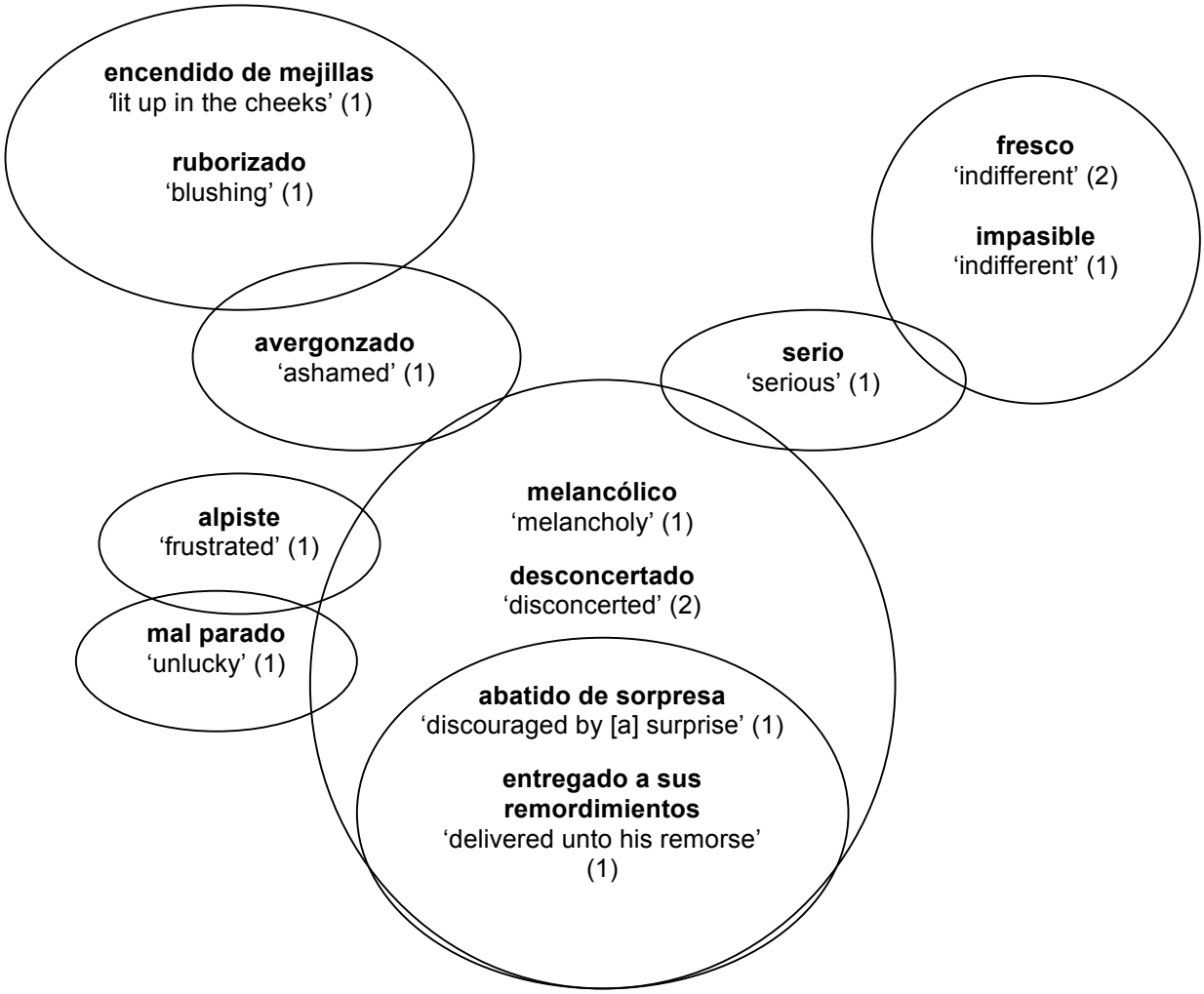


Figure 44, 1800's the *triste* clusters

3. The *confuso / suspenso* clusters

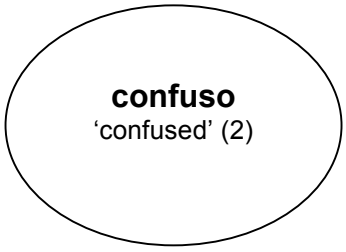


Figure 45, 1200's: The *confuso / suspenso* clusters.

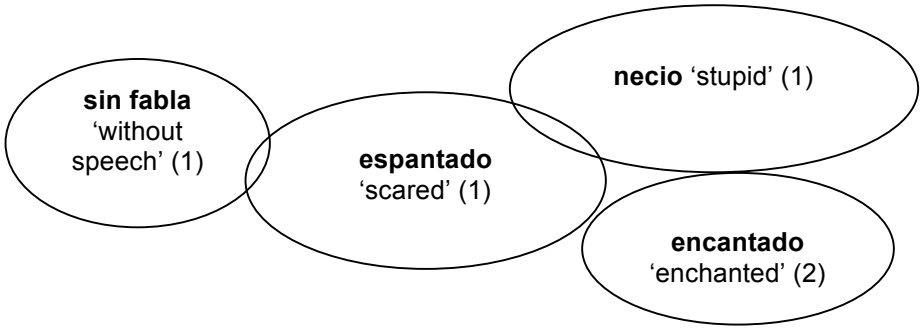


Figure 46, 1300's: The *confuso / suspenso* clusters.

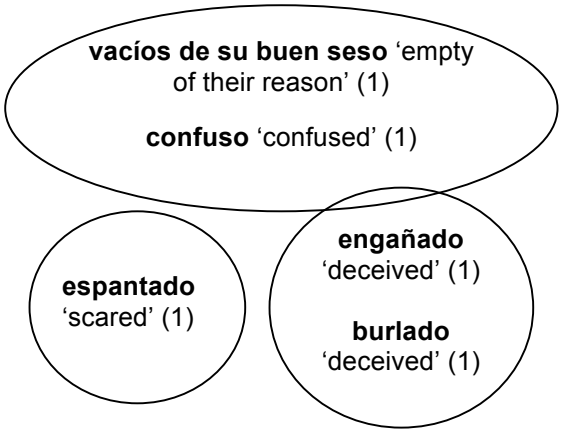


Figure 47, 1400's: The *confuso / suspenso* clusters.

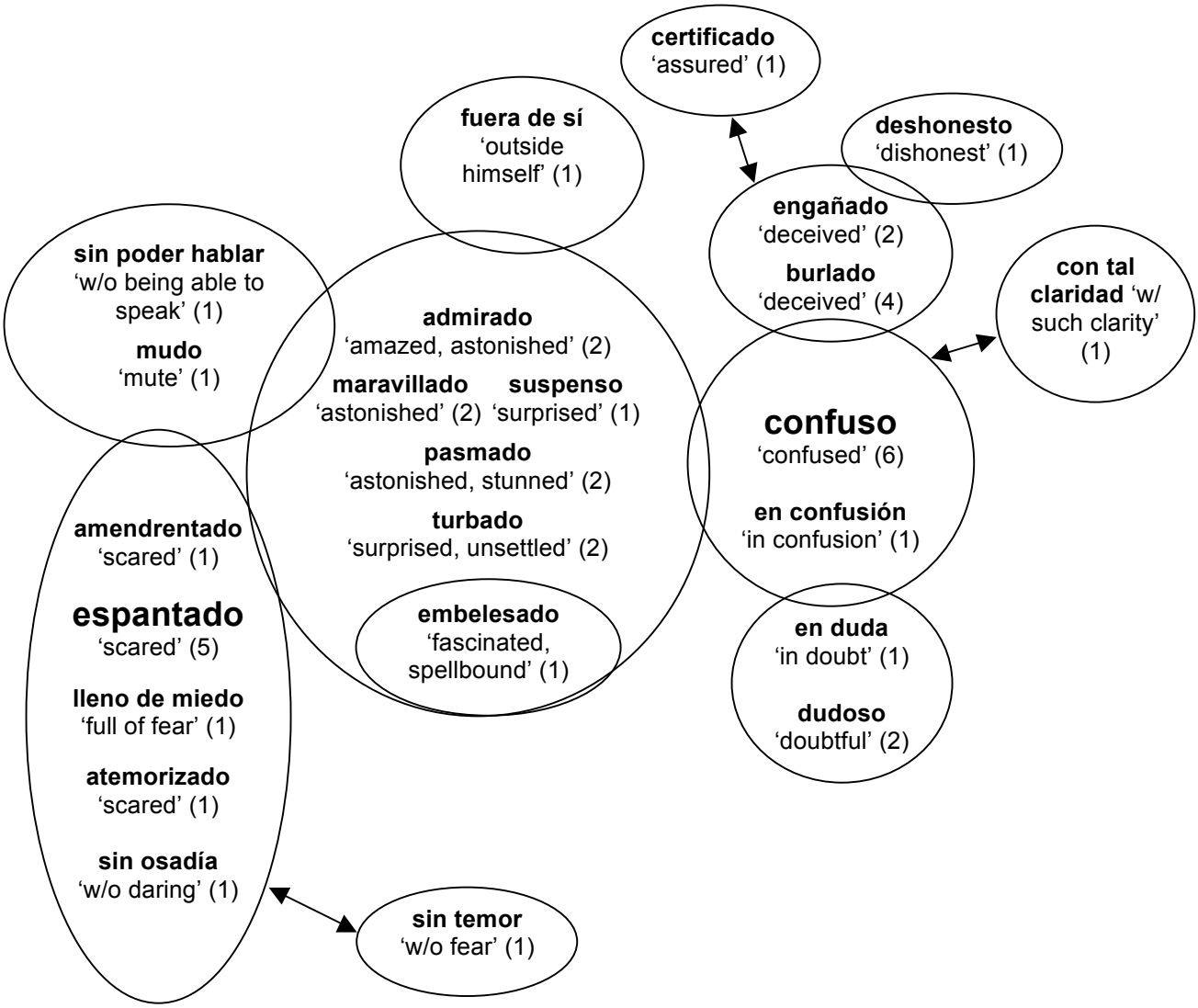


Figure 48, 1500's: The *confuso / suspenso* clusters.

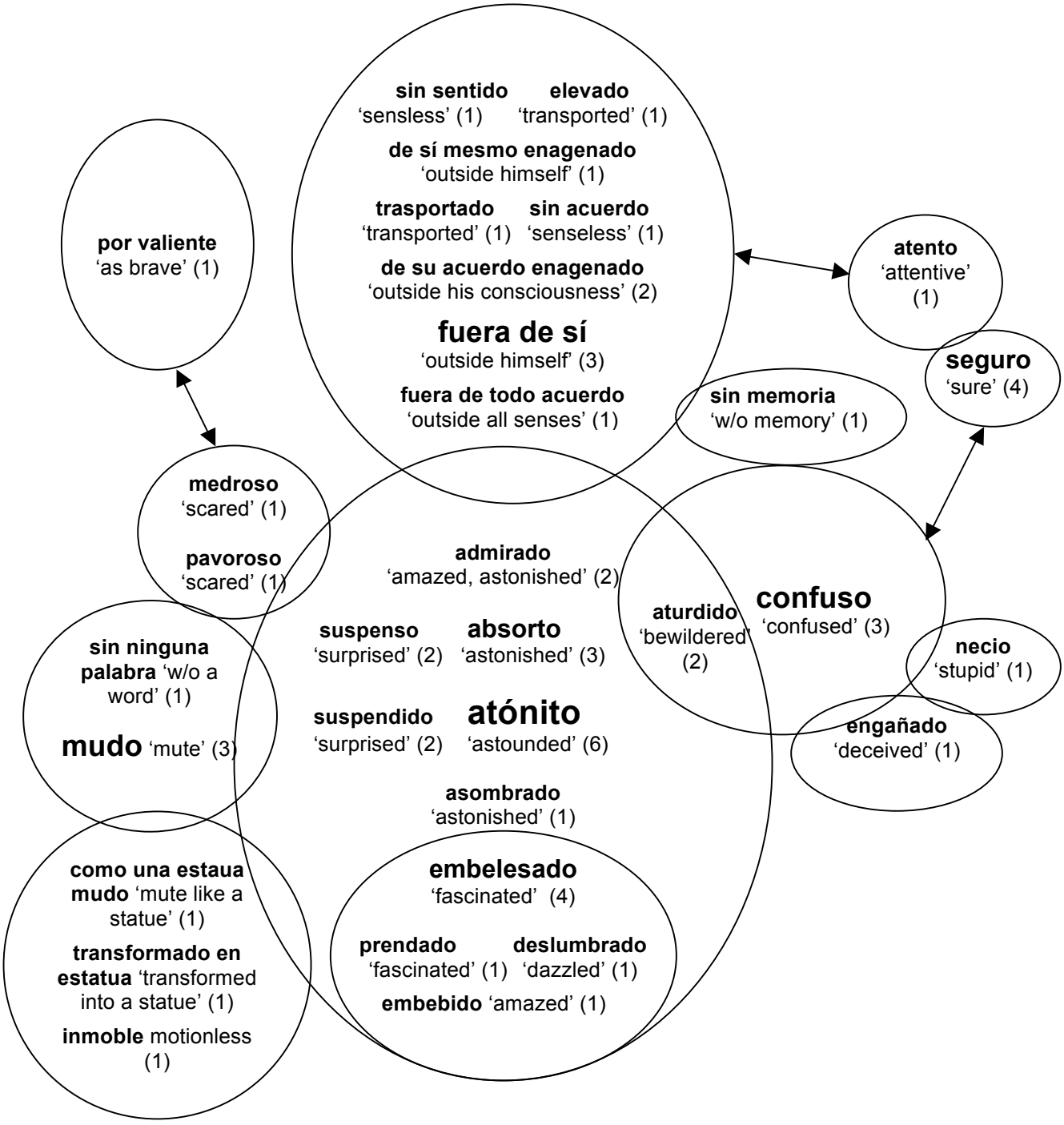


Figure 49, 1600's: The *confuso / suspenso* clusters.

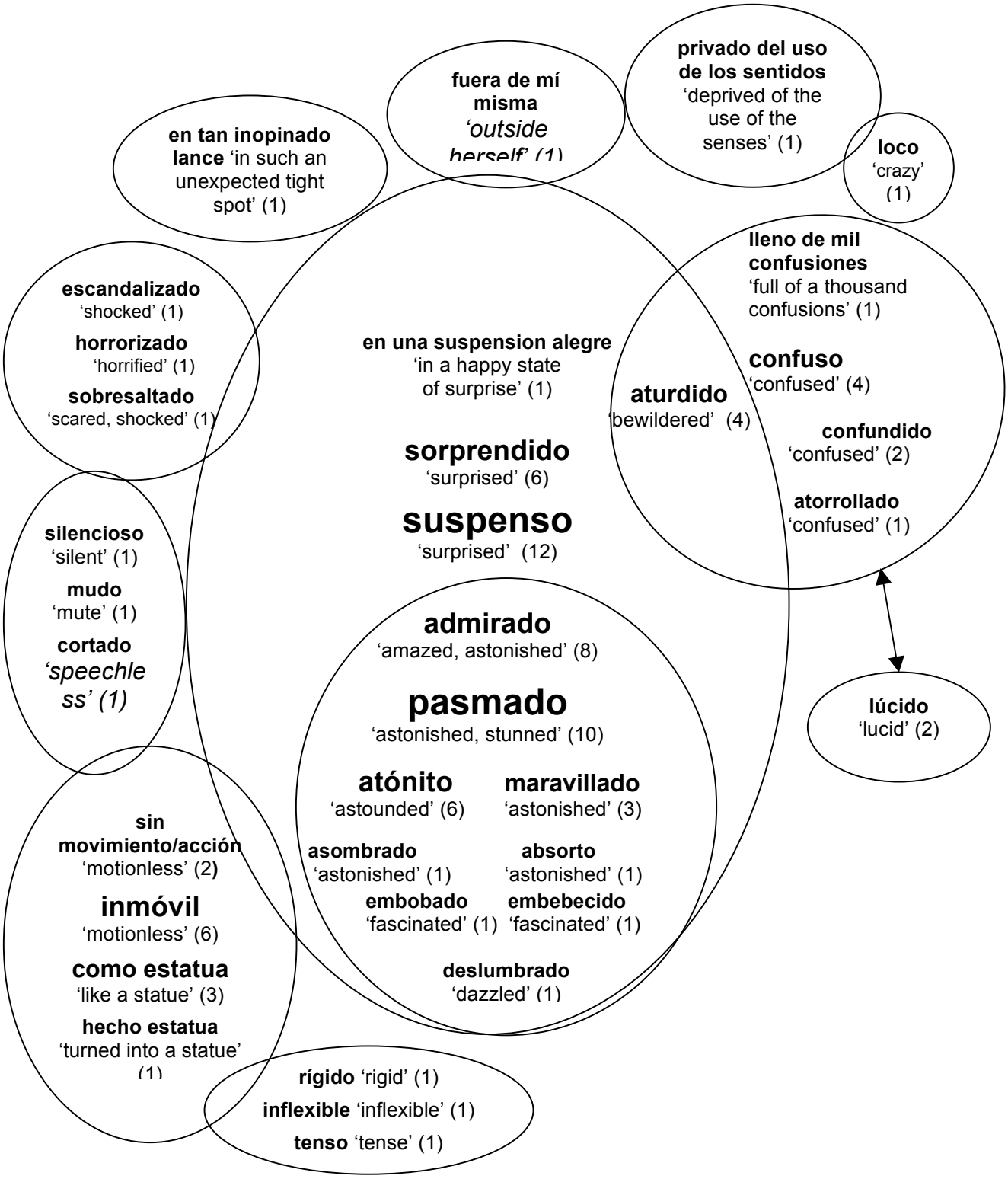


Figure 50, 1700's: The *confuso / suspenso* clusters.

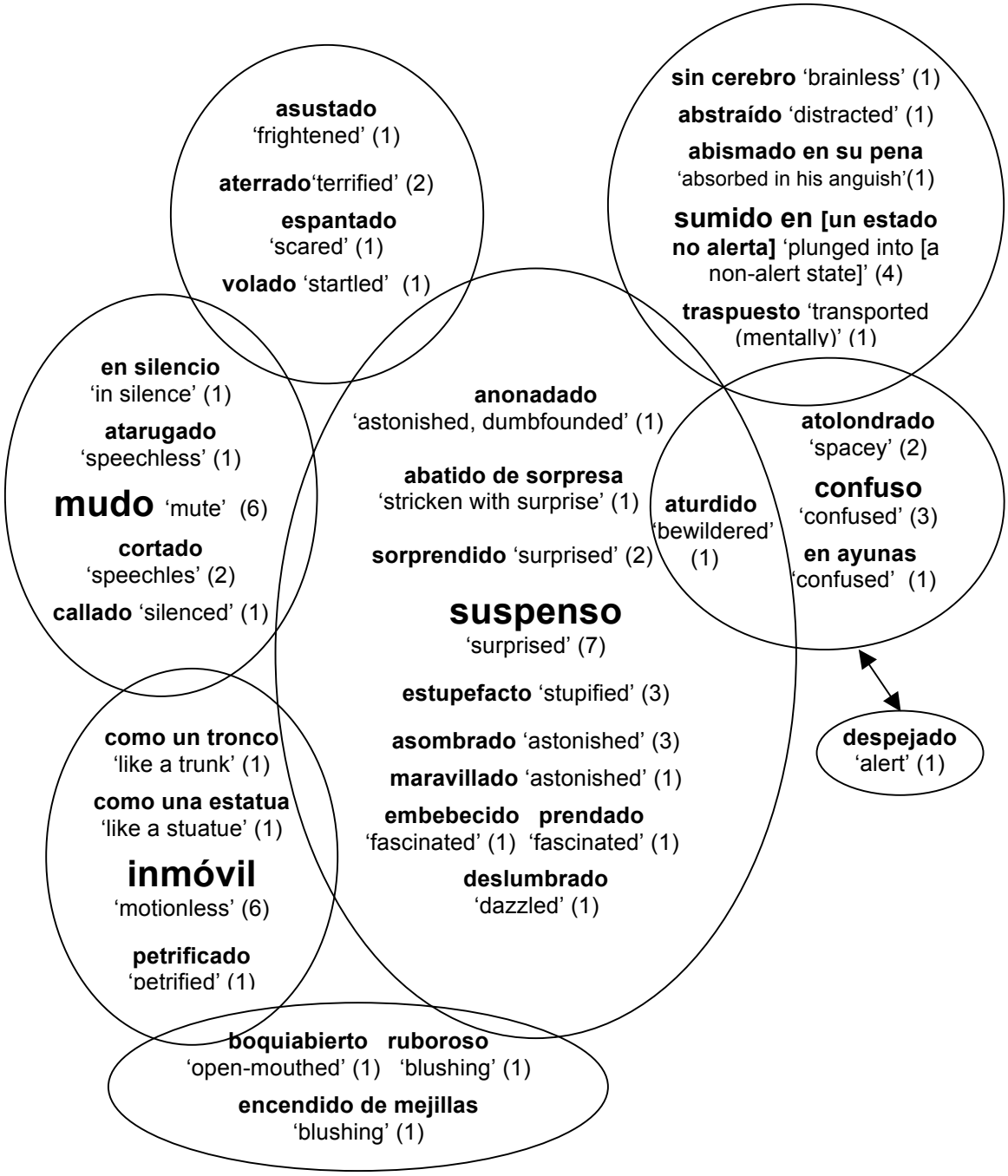


Figure 51, 1800's: The *confuso / suspenso* clusters.

4. The *convencido* clusters

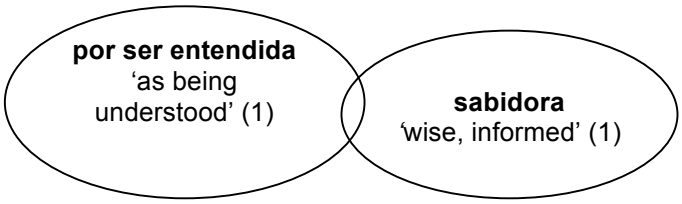


Figure 52, 1500's: the *convencido* clusters

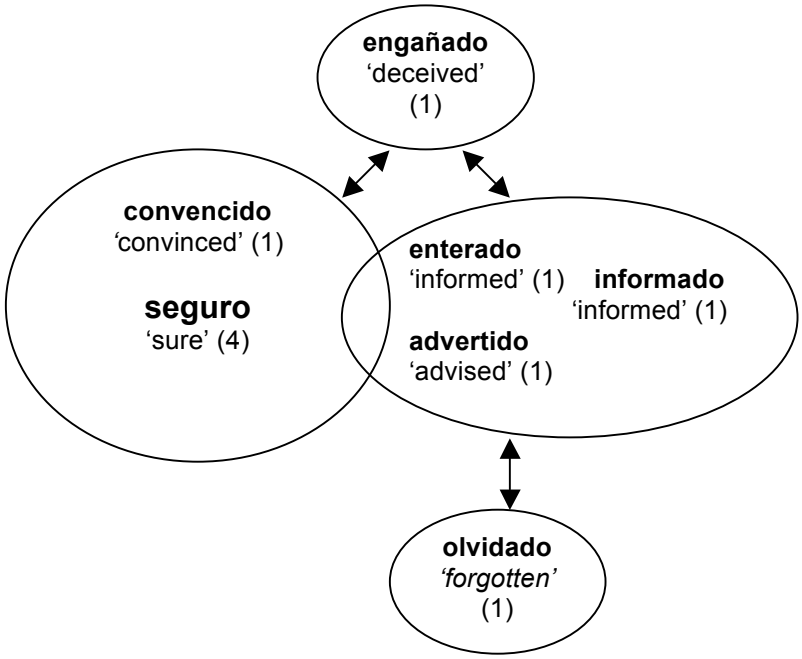


Figure 53, 1600's: the *convencido* clusters

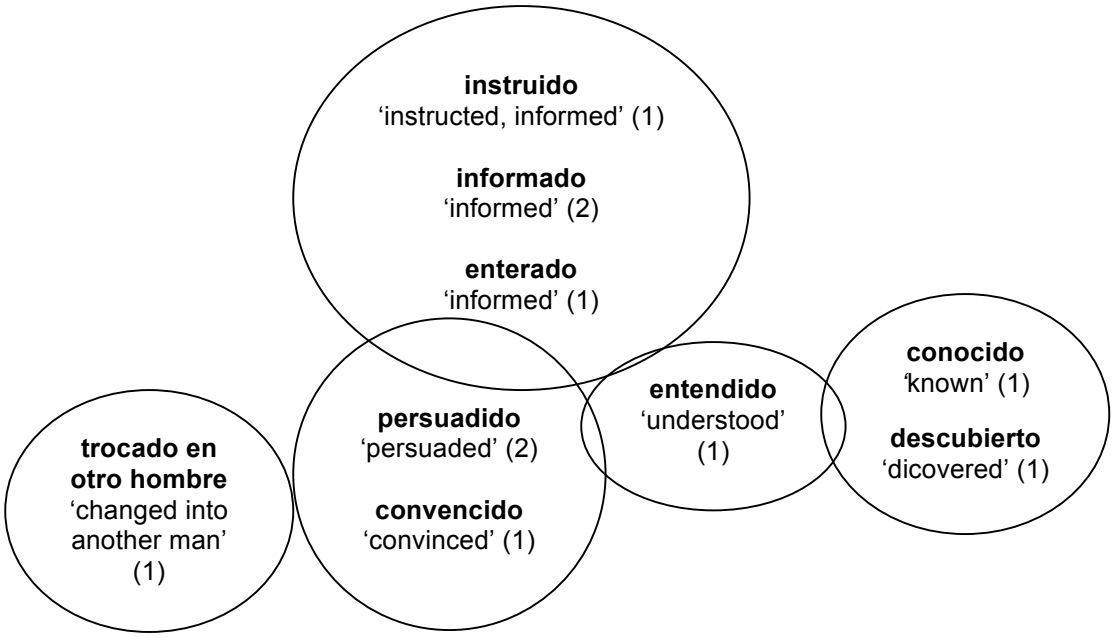


Figure 54, 1700's: the *convencido* clusters

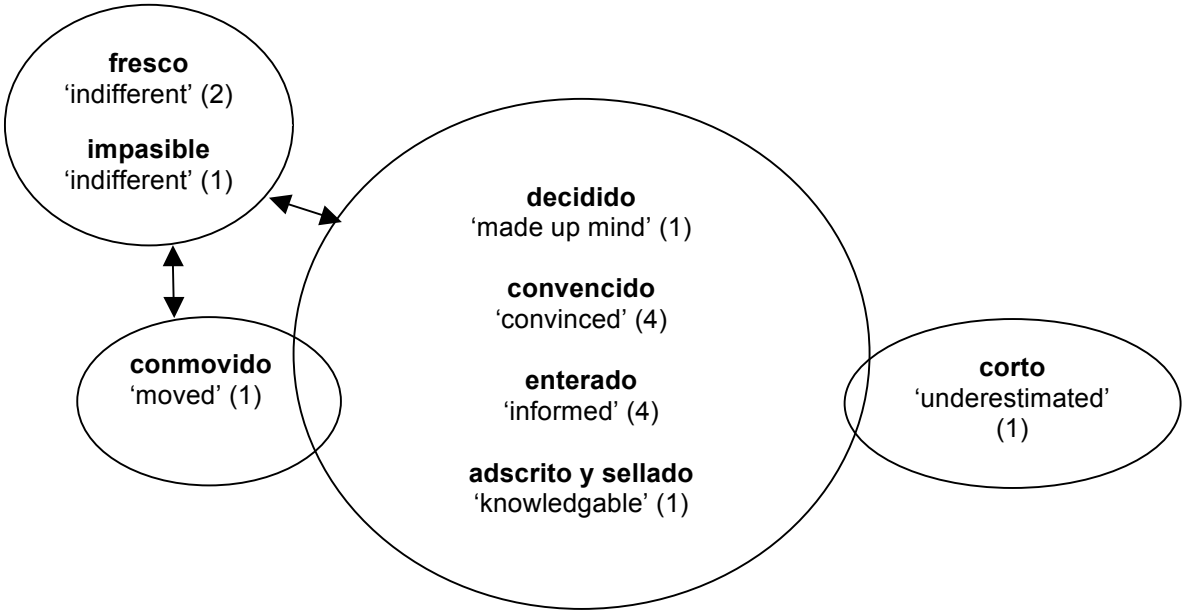


Figure 55, 1800's: the *convencido* clusters

5. Death, injury, and sometimes good health: the *muerto* clusters

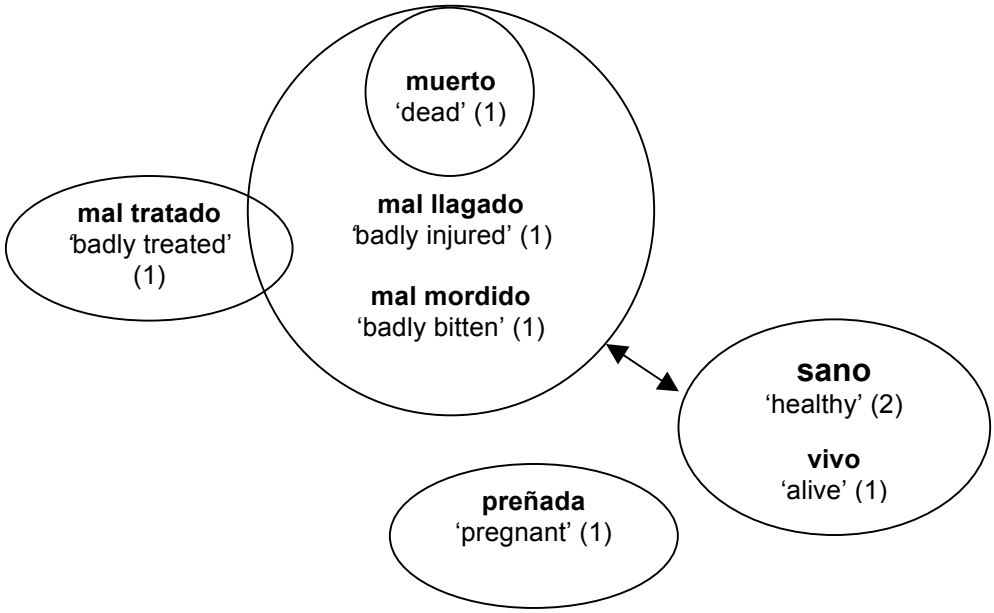


Figure 56, 1200's: the *muerto* clusters

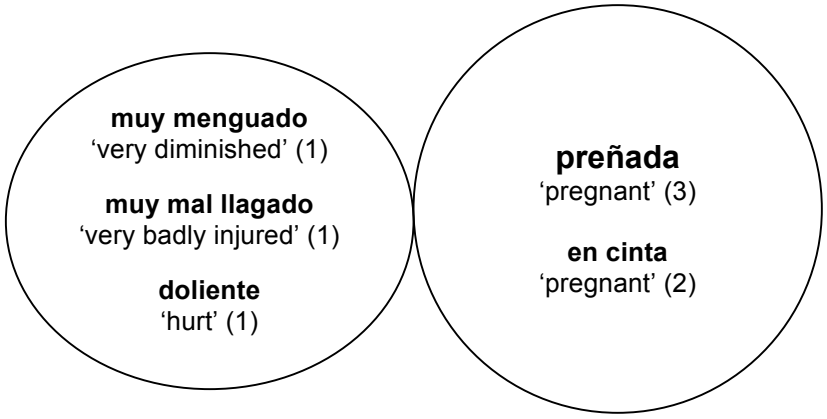


Figure 57, 1300's: the *muerto* clusters

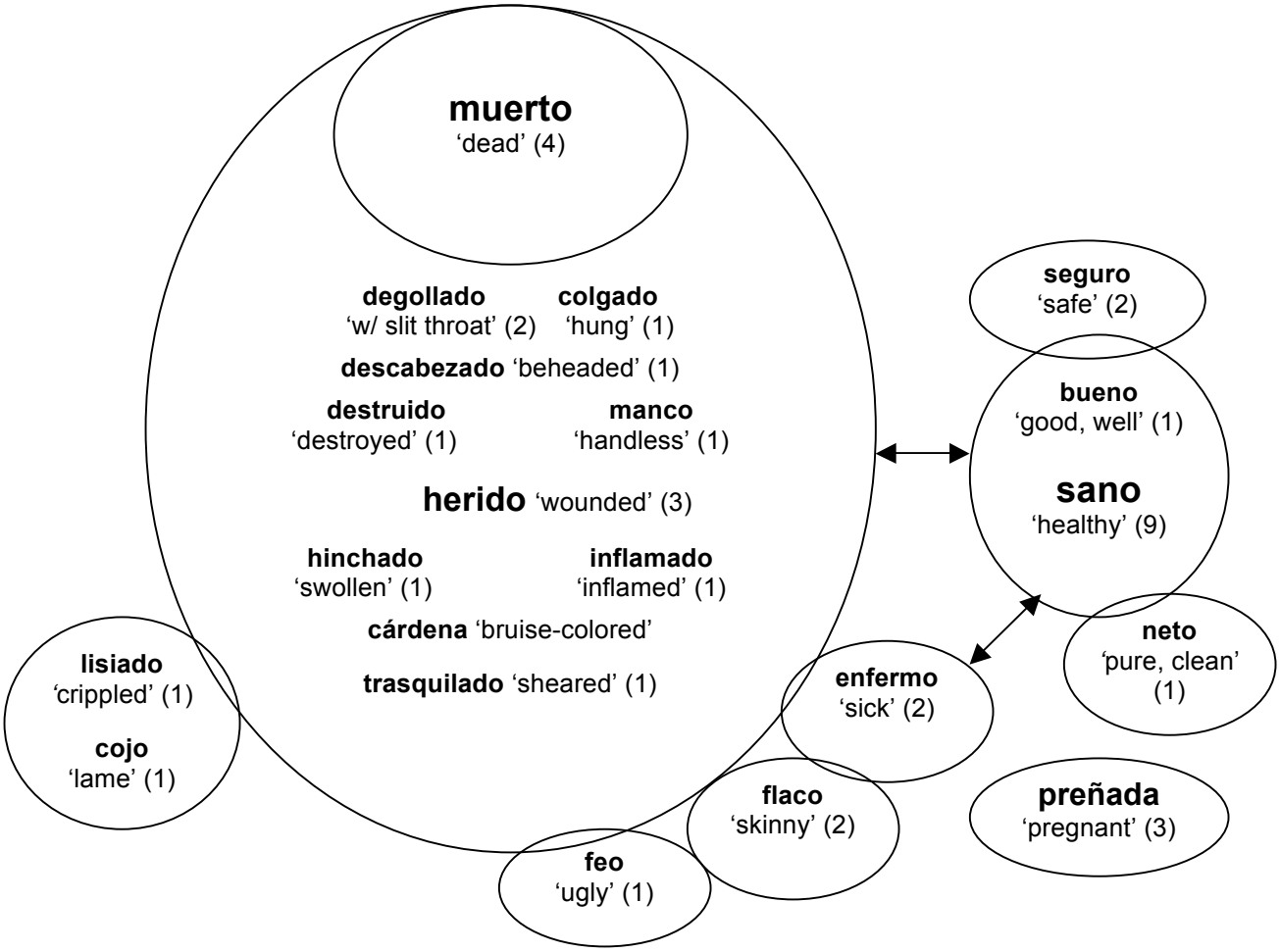


Figure 58, 1400's: the *muerto* clusters

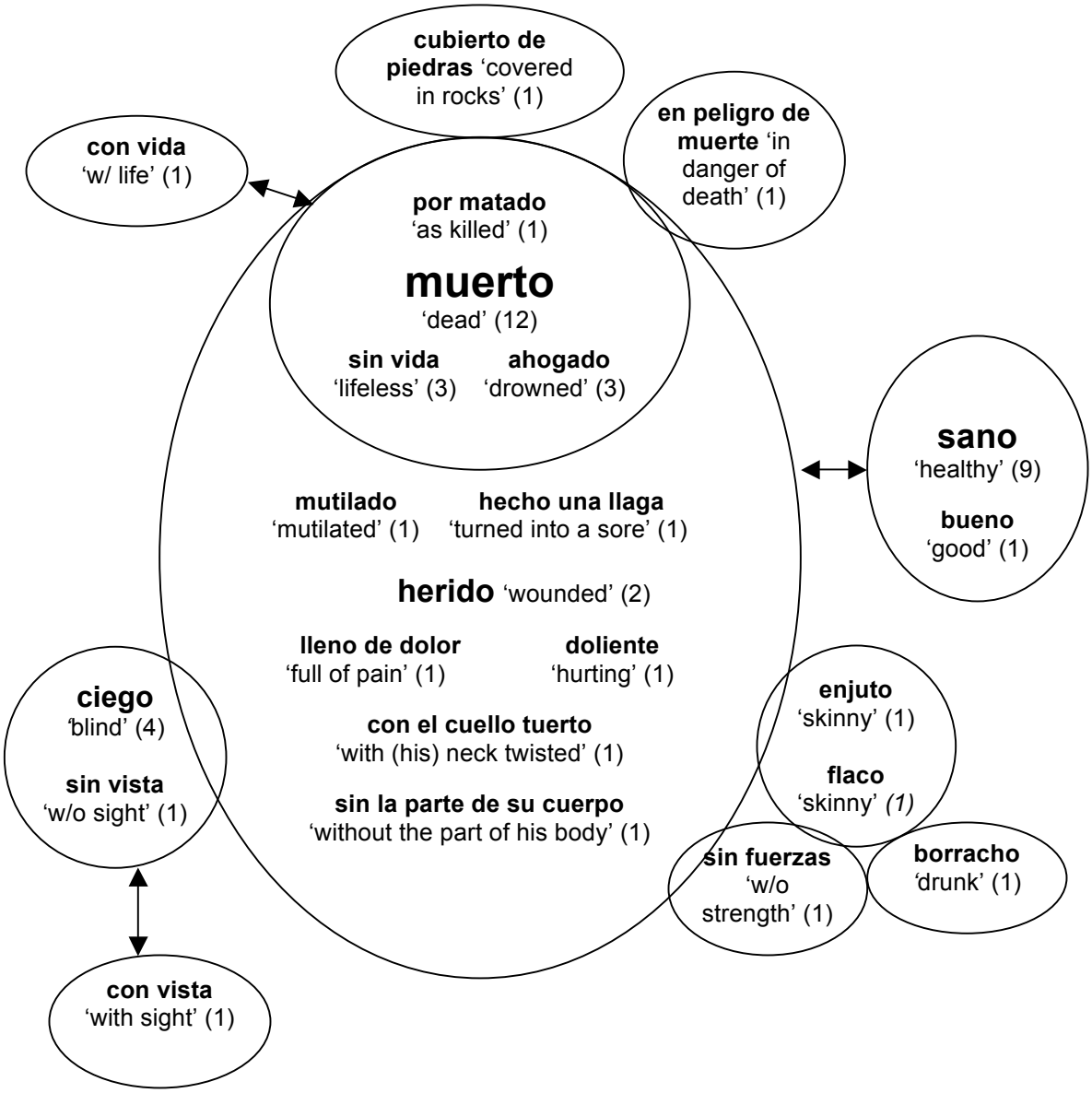


Figure 59, 1500's: the *muerto* clusters

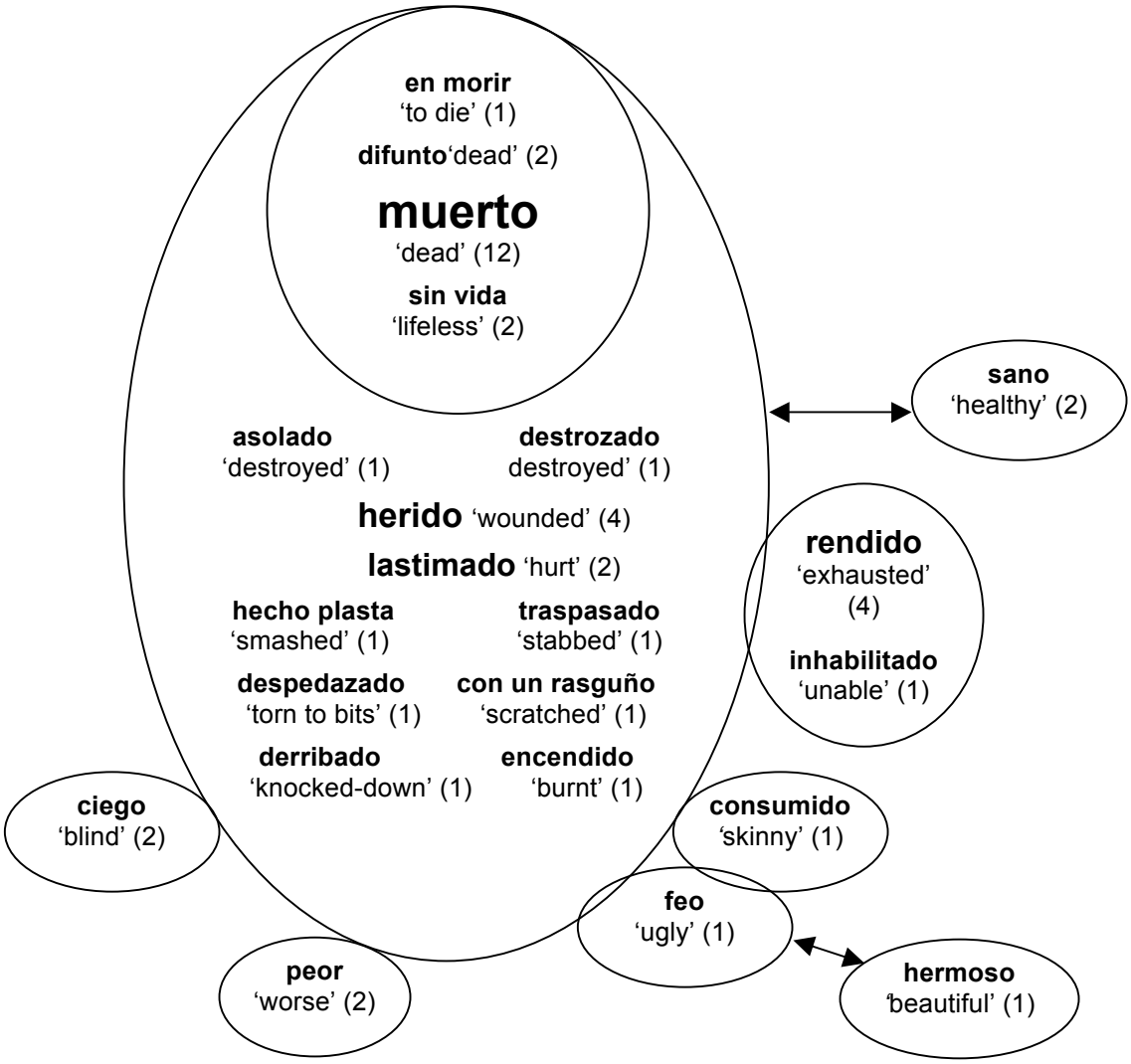


Figure 60, 1600's: the *muerto* clusters

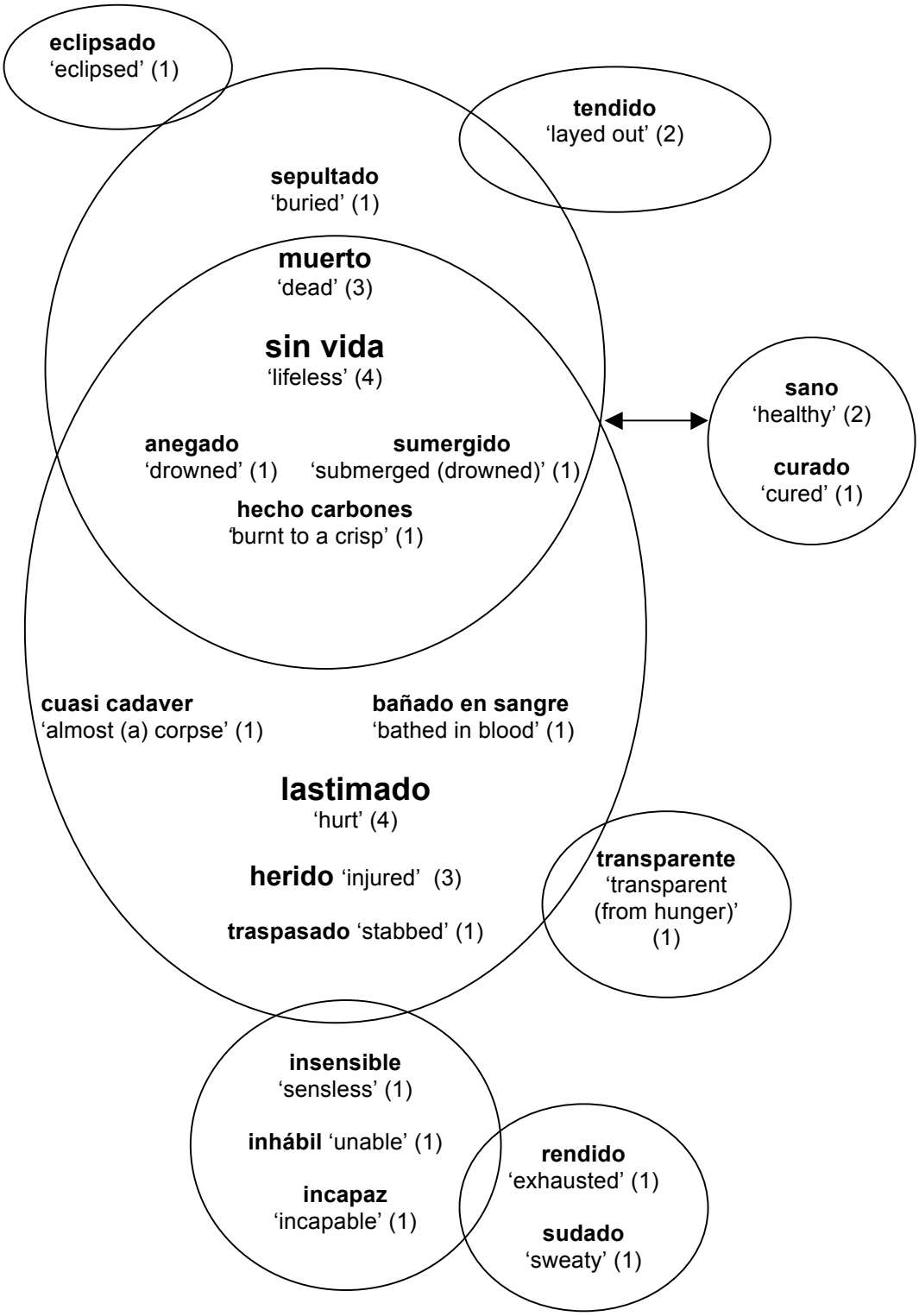


Figure 61, 1700's: the *muerto* clusters

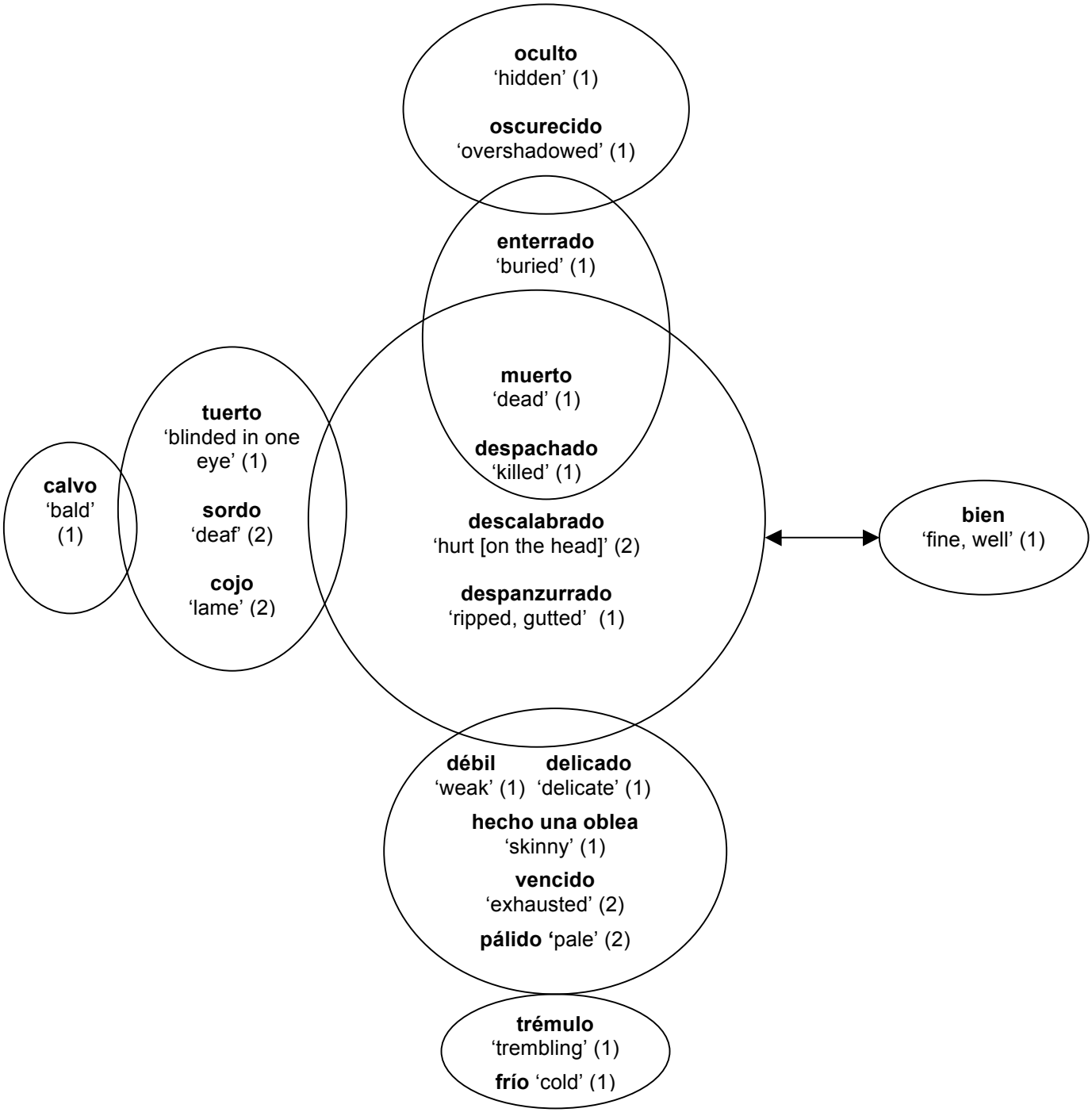


Figure 62, 1800's: the *muerto* clusters

6. Other physical states

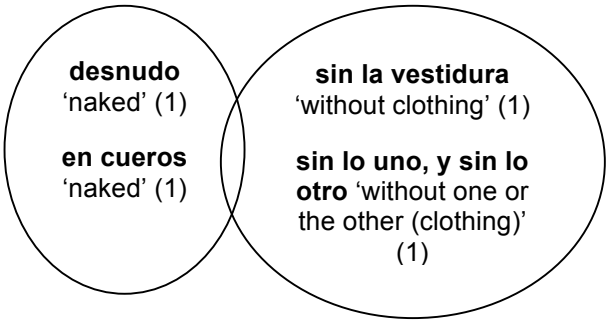


Figure 63, 1300's: nudity clusters

Table 36, 1800's Adjectives reflecting physical position

arrodillado *'kneeled'*	1
de rodillas *'kneeled'*	1
de pie *'on his/her feet'*	1
a pie *'on his/her feet'*	1
de espaldas *'back to back'*	1
sujeto *'stuck'*	1
enterrado *'buried'*	1
a oscuras *'in the dark'*	1
oculto *'hidden'*	
Conceptual:	
oscurecido *'overshadowed'*	1
Other:	
vestido *'dressed'*	1

Appendix 2: Sources used for data

Table 37. Works from the 1200's		
Title	Author	Source
Gran conquista de Ultramar.	Anónimo	Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings. Prepared by John ONeill. (Madison and New York, 1999). CD-ROM. (ISBN 1-56954-122-1). Salamanca Giesser 1503-06-21.
Castigos y documentos para bien vivir.	Sancho IV.	Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings. Prepared by John ONeill. (Madison and New York, 1999). CD-ROM. (ISBN 1-56954-122-1). Madrid Nacional ms. 6559, ff 1r-244v.
Los siete sabios de Roma	Anónimo	http://www.cervantesvirtual.com/FichaObra.html?Ref=0370
Poema de Fernán González	Anónimo	Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings. Prepared by John ONeill. (Madison and New York, 1999). CD-ROM. (ISBN 1-56954-122-1). Escorial Monasterio b-IV-21.

Table 38. Works from the 1300's		
Title	Author	Source
Biblia romanceada judio cristiana	Anónimo	Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings. Prepared by John ONeill. (Madison and New York, 1999). CD-ROM. (ISBN 1-56954-122-1). Madrid Real Academia Espanola codice 87.
Historia troyana	Guido de Columna	Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings. Prepared by John ONeill. (Madison and New York, 1999). CD-ROM. (ISBN 1-56954-122-1). Pamplona Arnaldo Guillén de Brocar 1499.
Leomarte: Sumas de la historia troyana. Madrid: Nacional MS. 9256.	Transcribed by Robert G. Black	Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings. Prepared by John ONeill. (Madison and New York, 1999). CD-ROM. (ISBN 1-56954-122-1). Madrid Nacional ms. 9256.
Crónica de 1344. Madrid: Zabalburu 11-109.	Transcribed by José P. da Cruz	Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings. Prepared by John ONeill. (Madison and New York, 1999). CD-ROM. (ISBN 1-56954-122-1). Madrid Nacional ms. 9256.

Table 39. Works from the 1400's		
Title	Author	Source
Triunfo de las donas y cadira de onor	Juan Rodríguez del Padrón	Cervantes Virtual
La Celestina	Fernando de Rojas ; edición y notas de Julio Cejador y Frauca	Cervantes Virtual
Libro de don Tristán de Leonís (Valladolid, 1501)	Anónimo	Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings; Prepared by John O'Neill. Madison & New York 1999
Castigos y doctrinas que un sabio daba a sus hijas.	Anónimo	Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings; Prepared by John O'Neill. Madison & New York 1999. Escorial: Monasterio a.IV.5. Transcribed by Connie L. Scarborough
Exemplario por ABC.	Anónimo	Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings; Prepared by John O'Neill. Madison & New York 1999. Madrid: Nacional MS. 1182. Transcribed by Javier Coca
Arcipreste de Talavera: Corbacho.		Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings; Prepared by John O'Neill. Madison & New York 1999, Escorial: Monasterio h.III.10. Transcribed by Eric W. Naylor
Atalaya de las corónicas.	Anónimo	Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings; Prepared by John O'Neill. Madison & New York 1999. London: British Library Egerton 287. Transcribed by James B. Larkin
Meditationes of Pseudo-Augustine (olim. De infantia Salvatoris).		Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings; Prepared by John O'Neill. Madison & New York 1999. Madrid: Nacional MS. Inc. 1424. Transcribed by Frank Waltman
Arnalte y Lucenda.	Anónimo	Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings; Prepared by John O'Neill. Madison & New York 1999. Milan: Trivulziana 940. Transcribed by Diane Wright
Cancionero de Baena	Anónimo	Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings; Prepared by John O'Neill. Madison & New York 1999. (Dutton PN1). Paris: Nationale Esp. 37. Transcribed by Victoria A. Burrus
Cancionero	Anónimo	Electronic Texts and Concordances of the Madison Corpus of Early

castellano y catalán de París		Spanish Manuscripts and Printings; Prepared by John O'Neill. Madison & New York 1999 .(Dutton PN4). Paris: Nationale Esp. 226. Transcribed by Robert G. Black
Boccacio: Caída de prinçipes (Seville, 1495)	Boccacio	Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings; Prepared by John O'Neill. Madison & New York 1999. New York: Hispanic Society. Transcribed by Eric Naylor
Claros varones de Castilla.	Anónimo	Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings; Prepared by John O'Neill. Madison & New York 1999. Madrid: Nacional I-1569. Transcribed by Michael L. Dangerfiel
Juan de Mena: Coronación.	Anónimo	Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings; Prepared by John O'Neill. Madison & New York 1999. New York: Hispanic Society. Transcribed by Ivy A. Corfis
Defensa de virtuossas mugeres.	Anónimo	Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings; Prepared by John O'Neill. Madison & New York 1999. Madrid: Nacional MS. 1341. Transcribed by María Isabel Montoya Ramírez
Generaciones y semblanzas.	Anónimo	Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings; Prepared by John O'Neill. Madison & New York 1999. Madrid: Fundación Lázaro Galdiano MS. 435. Transcribed by Robert Folger
Letras de Hernando del Pulgar (Burgos, 1485).	Hernando del Pulgar	Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings; Prepared by John O'Neill. Madison & New York 1999. Transcribed by Ivy A. Corfis
Historia de la linda Melosina.	Anónimo	Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings; Prepared by John O'Neill. Madison & New York 1999. London: British Library IB.42463. Transcribed by Ivy A. Corfis
Morales de Ovidio.	Anónimo	Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings; Prepared by John O'Neill. Madison & New York 1999. Madrid: Nacional MS. 10144. Transcribed by Derek Carr
Oliveros de Castilla (Burgos, 1499).	Anónimo	Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings; Prepared by John O'Neill. Madison & New York 1999. New York: Hispanic Society. Transcribed by Ivy A. Corfis
Suma de las corónicas.	Anónimo	Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings; Prepared by John O'Neill. Madison & New York 1999. Escorial: Monasterio h.II.22. Transcribed by Judy Krieger
Historia del gran Tamerlán.	Anónimo	Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings; Prepared by John O'Neill. Madison & New York 1999. Madrid: Nacional MS. 9218. Transcribed by Juan Luis Rodríguez Bravo and María del Mar Martínez Rodríguez
Las Siete	Pablo de	Textos Lemir. Edición y estudio a cargo de Juan Carlos Conde 1997

edades del mundo	Santa María. Refundición de 1460	
De las mujeres illustres en romance, Caps. i-xxx	Johan Boccaccio, 1494	Textos Lemir. 1997, José Luis Canet

Table 40. Works from the 1500's		
Title	Author	Source
COMEDIA LLAMADA ROSABELLA	Martín de Santander (1550)	Lemir
Comedia llamada Florinea	Juan Rodríguez Florián, 1554	Lemir
La Comedia de Anfitrión	de Juan Timoneda (1559)	Lemir. Ed. de Manuel V. Diago (© 2000)
La Comedia de los Menennos	de Juan Timoneda (1559)	Lemir. Ed. de Manuel V. Diago (© 2000)
La Comedia Cornelia o Carmelia	de Juan Timoneda (1559)	Lemir. Ed. de Manuel V. Diago (© 2000)
COMEDIA VIDRIANA	de JAIME DE HUETE (1525?)	Lemir. Edición realizada por Coronada Requena Pineda (Universitat de València)
FARSA DEL MUNDO Y MORAL	de Fernán López de Yanguas (1524)	Lemir. Edición realizada por Ana Belén Esteve López (Universidad de Valencia) 10/07/00
Farsa del Sordo, (Valladolid, s.a; y Burgos, 1561)	ed. de Gillermo Soler (© 2002)	Lemir. (ed. electrónica, José L. Canet)
Comedia Tibalda	de Perálvarez de Ayllón (1553)	Lemir. ed. de Vicent Caparrós (© 2002)
Aula de cortesanos	Cristobal de Castillejo (1494-1550)	Cervantes Virtual.

Auto de la Pasión	Lucas Fernández	Cervantes Virtual. Edición digital a partir de la edición facsímil de la de Salamanca, 1514, realizada por Emilio Cotarelo y Mori (Madrid, Real Academia Española, 1929) y cotejada con la edición crítica de Mª Josefa Canellada (Madrid, Castalia, 1976, pp. 211-237).]
Diálogo de las cosas acaecidas en Roma	Alfonso de Valdés, (1530?)	Antologia de ensayo. ha sido preparada por la Biblioteca Virtual Miguel de Cervantes
Diálogo de Mercurio y Carón	Alfonso de Valdés, (1530?)	Antologia de ensayo. ha sido preparada por la Biblioteca Virtual Miguel de Cervantes
Tragicomedia llamada Filomena	Joan Timoneda (1564)	Lemir. 2003) Rómulo Pianacci (editor) Ed. electrónica José L. Canet
La penitencia del amor	de Pedro Manuel de Urrea (1514)	Lemir. ed. de José Luis Canet (© 2003)
La Comedia Ypólita [Hipólita]	Anónimo valenciano de (1521)	Lemir. ed. de José Luis Canet (© 2003)
La Comedia Serafina	Anónimo valenciano de (1521)	Lemir. ed. de José Luis Canet (© 2003)
La Comedia Grassandora	Juan Uceda de Sepúlveda (1540)	Lemir. ed. de José Luis Canet (© 2003)
Farsa a manera de tragedia	(Valencia, 1537)	Lemir. ed. de Isabel Pascual Lavilla (© 2003)
El Coloquio de Camilla	Lope de Rueda, (1567)	Lemir. ed. de Inmaculada Arlandis (© 2005)
Castigos y enxemplos de Catón	Medina del Campo 1543	Lemir.
Laberinto de amor	(1546) de Juan Boccaccio.	Lemir. Edición realizada por Diego Romero Lucas.
Sobremesa y alivio de caminantes	Joan Timoneda, (Valencia, 1569)	Lemir.
Fructus Sanctorum y Quinta Parte del Flos Sanctorum ,	de Alonso de Villegas (1594)	Lemir. (c)1998 Ed. de José Aragüés Aldaz

Table 41. Works from the 1600's		
Title	Author	Source
Aventurarse perdiendo	Zayas y Sotomayor, María de (1590-¿1661?)	Alicante : Biblioteca Virtual Miguel de Cervantes, 1999 Nota: Edición digital a partir de la edición de Augstín González de Amezúa, Novelas amorosas y ejemplares, Madrid, Real Academia Española, 1948 (Biblioteca Selecta de Clásicos Españoles. Serie II ; 8).
Baile de los oficios	Moreto, Agustín (1618-1669)	licante : Biblioteca Virtual Miguel de Cervantes, 1999 Nota: Edición digital basada en la de Madrid, Biblioteca de Autores Cristianos, 1971
El castigo de la miseria	Zayas y Sotomayor, María de (1590-¿1661?)	Alicante : Biblioteca Virtual Miguel de Cervantes, 1999 Nota: Edición digital a partir de la edición de Augstín González de Amezúa, Novelas amorosas y ejemplares, Madrid, Real Academia Española, 1948 (Biblioteca Selecta de Clásicos Españoles. Serie II ; 8)
La esclava de su amante	Zayas y Sotomayor, María de (1590-¿1661?)	Alicante : Biblioteca Virtual Miguel de Cervantes, 1999 Nota: Edición digital a partir de la edición de Augstín González de Amezúa, Novelas amorosas y ejemplares, Madrid, Real Academia Española, 1948 (Biblioteca Selecta de Clásicos Españoles. Serie II ; 8).
Hospital de incurables. Viaje de este mundo y el otro	Jacinto Polo de Medina	Alicante : Biblioteca Virtual Miguel de Cervantes, 2002 Nota: Edición digital a partir de la de Orihuela, Juan Vicente Franco, 1636.
La inocencia castigada & El jardín engañoso	Zayas y Sotomayor, María de (1590-¿1661?)	Alicante : Biblioteca Virtual Miguel de Cervantes, 1999 Nota: Edición digital a partir de la edición de Augstín González de Amezúa, Novelas amorosas y ejemplares, Madrid, Real Academia Española, 1948 (Biblioteca Selecta de Clásicos Españoles. Serie II ; 8).
Abre el ojo/ de Francisco de Rojas Zorrilla ; ordenadas en colección por Ramón de Mesonero Romanos	Rojas Zorrilla, Francisco de (1607-1648)	Mesonero Romanos. Editor Nota: Edición digital a partir de Comedias escogidas de D. Francisco de Rojas Zorrilla, Madrid, Imprenta de los Sucesores de Hernando, 1918.
Afectos de odio y amor	Calderón de la Barca, Pedro (1600-1681).	Alicante : Biblioteca Virtual Miguel de Cervantes, 2002. Nota: Edición digital a partir de la Tercera parte de Comedias de don Pedro Calderón de la Barca, Madrid, Por Domingo García Morràs, a costa de Domingo Palacio y Villegas, 1664.
El alcalde de Zalamea / Pedro Calderón de la Barca ; edición de José María	Calderón de la Barca, Pedro (1600-1681).	Alicante : Biblioteca Virtual Miguel de Cervantes, 1999 Nota: Otra ed.: José María Ruano de la Haza, Madrid, Espasa Calpe, 1995, 5ª ed.

Ruano de la Haza		
Los amantes de Teruel	Molina, Tirso de (¿1583?-1648)	Alicante : Biblioteca Virtual Miguel de Cervantes, 2006 Nota: Edición digital a partir de Segvnda parte de las comedias... recogidas por sv sobrino don Francisco Lucas de Auila, Madrid, en la imprenta del Reino, a costa de la hermandad de los Mercaderes de Libros desta Corte, 1635. Localización : Base de Datos Teatro Español del Siglo de Oro (TESO).
Amar sin saber a quién	Vega, Lope de (1562-1635)	Alicante : Biblioteca Virtual Miguel de Cervantes, 2002 Nota: Edición digital a partir de Ventidos parte perfeta de las comedias del Fenix de España Frey Lope Felix de Vega Carpio... Madrid, por la viuda de Juan Gonzalez, a costa de Domingo de Palacio y Villegas y Pedro Verges..., 1635. Localización: Base de Datos Teatro Español del Siglo de Oro (TESO). Autorizada por Miguel Ángel Auladell Pérez.
Amazonas en las Indias	Vega, Lope de (1562-1635)	Alicante : Biblioteca Virtual Miguel de Cervantes, 1999 Nota: Edición digital a partir de la de Quarta parte de las comedias del Maestro Tirso de Molina, Madrid, María de Quiñones, 1635.
La amistad castigada	Ruiz de Alarcón y Mendoza, Juan (1581-1639)	Alicante : Biblioteca Virtual Miguel de Cervantes, 1999 Nota: Edición digital a partir de Obras Completas. T. I, Valencia, Albatros Hispanofila, 1990, pp. 69-100
El amor constante	Castro, Guillén de (1569-1631)	Alicante : Biblioteca Virtual Miguel de Cervantes, 2006 Nota: Edición original en la Biblioteca de la Universidad de Alicante Nota: Edición digital a partir de Obras Completas, Edición y prólogo de Joan Oleza, Madrid, Fundación José Antonio de Castro, 1997, pp. 4-115
EL LABERINTO DE AMOR	Miguel de Cervantes Saavedra	Texto basado en la edición príncipe, EL LABERINTO DE AMOR en OCHO COMEDIAS Y OCHO ENTREMESES NUEVOS NUNCA REPRESENTADOS, COMPUESTAS POR MIGUEL DE CERVANTES SAAVEDRA (Madrid: Viuda de Alonso Martín, 1615). Fue editado en forma electrónica por Vern G. Williamsen en 1997.
LA NUMANCIA	Miguel de Cervantes Saavedra	Texto preparado en 1996 por Vern G. Williamsen con el apoyo de la edición de José Martel y Hymen Alpern en su DIEZ COMEDIAS DEL SIGLO DE ORO (New York: Harper and Row, 1939).
EL RUFIÁN DICHOSO & El retablo de las maravillas	Miguel de Cervantes Saavedra	Texto basado en la edición príncipe, EL LABERINTO DE AMOR en OCHO COMEDIAS Y OCHO ENTREMESES NUEVOS NUNCA REPRESENTADOS, COMPUESTAS POR MIGUEL DE CERVANTES SAAVEDRA (Madrid: Viuda de Alonso Martín, 1615). Fue editado en forma electrónica por Vern G. Williamsen en 1997.
LA ADVERSA FORTUNA DE DON BERNARDO DE CABRERA	Antonio Mira de Amescua	Texto basado en varios textos tempranos del LA ADVERSA FORTUNA DE DON BERNARDO DE CABRERA. Fue preparado por Vern Williamsen en el año 1976. El texto base para esta edición es el príncipe, Doce comedias de Lope de Vega Carpio y otros autores, parte veinte y nueve (Huesca: Pedro Blusón, 1634). Luego fue editado en forma electrónica en el año 1986.
EL EJEMPLO	Antonio	Texto basado en el autógrafo de EL EJEMPLO MAYOR DE LA

MAYOR DE LA DESDICHA	Mira de Amescua	DESDICHA (Biblioteca Nacional, Madrid, R-112) con el apoyo de la edición príncipe, Parte veinticinco de comedias recopiladas de diferentes autores e ilustres poetas de España (Zaragoza: Pedro Escuer, 1632). Esta edición fue preparada por Vern Williamsen para un curso dictado en el año 1984.
EL DESDÉN CON EL DESDÉN	Agustín Moreto	Texto basado en varios textos tempranos del EL DESDÉN CON EL DESDÉN. Fue preparado en forma electrónica por Vern Williamsen en el año 1995.
EL HIJO DEL SERAFÍN, SAN PEDRO DE ALCÁNTARA	Juan Pérez de Montalbán	Texto basado en el texto del PRIMER TOMO DE LAS COMEDIAS DEL DO[C]TOR JUAN PÉREZ DE MONTALBÁN (Valencia: Claudio Macé, 1652).
Los dos soles de Toledo	Alonso Alcalá y Herrera	Edición digital a partir de Varios efectos de Amor en cinco novelas ejemplares, en Lisboa, por Manuel de Silva, 1641 y cotejada con la edición crítica de Evangelina Rodríguez, Novelas amorosas de diversos ingenios del siglo XVII, Madrid, Castalia, 1986, pp. 203-231.]
Los efectos de la fuerza	José Camerino	Edición digital a partir de Novelas amorosas..., en Madrid, por Tomás Iunti, 1624, pp. 159-174 y cotejada con la edición crítica de Evangelina Rodríguez, Novelas amorosas de diversos ingenios del siglo XVII, Madrid, Castalia, 1986, pp. 109-126.
La fantasma de Valencia	Alonso de Castillo Solórzano	Edición digital a partir de Tardes entretenidas, en Madrid, por viuda de Alonso Martín, 1625, pp. 54-85 y cotejada con la edición crítica de Evangelina Rodríguez, Novelas amorosas de diversos ingenios del siglo XVII, Madrid, Castalia, 1986, pp. 169-200.
La Constante Amarilis	Cristóbal Suárez de Figueroa (Valencia, 1609)	Textos Lemir. (2002) ed. de María Asunción Satorre Grau
Sátira de las comedias que escribió Mosén Pedro Morlá contra el sermón que predicó Don Luis Crespí	Mosén Pedro Morlá (Valencia, c.1649)	Textos Lemir. Introducción y edición ABRAHAM MADROÑAL
GENEALOGÍA DE LA TOLEDANA DISCRETA (Cantos I-XII)	Eugenio Martínez (Toledo, 1604)	Textos Lemir. edición y notas de Juan Carlos Pantoja Rivero

Table 42. Works from the 1700's		
Title	Author	Source
Diálogos nuevos en español y en francés,	Francisco Sobrino, 1708	Antología del Ensayo Ibero e Iberoamericano. Ed. de Daniel M. Sáez Rivera © (2002)

"Voz del Pueblo"	Benito Jerónimo Feijoo, 1726	Biblioteca Virtual Miguel de Cervantes. José Luis Gómez-Martínez
Los desahuciados del mundo y de la gloria	Diego de Torres Villarroel, (1736)	Biblioteca Virtual Miguel de Cervantes. Imp. A. Villagordo y P. Ortiz Gómez, 1752 y cotejada con la excelente edición de Manuel Mª Pérez López (Madrid, Editora Nacional, 1979).
El Valdemaro	Vicente Martínez Colomer 1792 OP	Biblioteca Virtual Miguel de Cervantes. Alicante : Biblioteca Virtual Miguel de Cervantes, 2002 Nota: Edición digital basada en la 4ª ed. de Valencia, José Ferrer de Orga, 1816
Los trabajos de Narciso y Filomela : (una novela cervantina del siglo XVIII)	Vicente Martínez Colomer	Biblioteca Virtual Miguel de Cervantes. Edición de Antonio Cruz Casado.
Fray Gerundio de Campazas (Intro y Libros I-III)	José Francisco de Isla (1758)	Biblioteca Virtual Miguel de Cervantes. Edición digital a partir de la de Madrid, Imprenta de Gabriel Ramírez, 1758 y cotejada con la edición de Russell P. Sebold (Madrid, Espasa Calpe, 1992, 3.ª ed.
Discursos forenses	Juan Meléndez Valdés (1798 ish)	Biblioteca Virtual Miguel de Cervantes. Edición de Emilio Palacios Fernández
El atolondrado : pieza original en un acto	Don Vicente Rodríguez Arellano (1750-1806)	Biblioteca Virtual Miguel de Cervantes.
Las bodas de Camacho el Rico	Dr. D. Juan Meléndez Valdés (1784)	Biblioteca Virtual Miguel de Cervantes. Edición digital basada en la edición de Madrid, por Joachin Ibarra, Impresor de Cámara de S.M., 1784.
El celoso don Lesmes : comedia nueva en tres actos	D.Vicente Rodríguez de Arellano (1790 (1750-1806))	Biblioteca Virtual Miguel de Cervantes. Edición digital basada en la edición de [Madrid], se hallará en la Librería de Castillo, ca 1790. Localización: Biblioteca de la Universidad de Castilla - La Mancha.
El calderero de San Germán, o El mutuo agradecimiento : comedia nueva en tres actos	Don Gaspar Zavala y Zamora (1762-¿1824? (ca.1790))	Antología del Ensayo Ibero e Iberoamericano. Edición digital basada en la edición de [Madrid], se hallará en la Librería Cerro, ca.1790.Localización: Biblioteca de la Universidad de Castilla - La Mancha

Table 43. Works from the 1800's		
Title	Author	Source
Aben Humeya o La rebelión de los moriscos,	D. Francisco Martínez de la Rosa	Obras literarias de D. Francisco Martínez de la Rosa, París, Imprenta de Julio Didot, 1827-30, t. V, 1830, pp. 119-240, y cotejada con la edición de Jean Sarrailh, Madrid, Espasa-Calpe, 1954.
Acertar errando o El cambio de diligencia : comedia en tres actos	Vega, Ventura de la (1807-1865)	Vega, Ventura de la (1807-1865), Alicante : Biblioteca Virtual Miguel de Cervantes, 2000. Edición digital basada en la edición de Madrid, Imprenta de Repullés, 1832. Localización: Biblioteca de la Universidad de Castilla La Mancha (Ciudad Real).
El afán de figurar, Comedia en cinco actos, en verso, acomodada al teatro español	José María de Carnerero	Cervantes Virtual
Los africanistas :	Gabriel Merino y Enrique López Marín	humorada cómico-lírica en un acto, dividido en tres cuadros (consecuencia de "El dúo de la Africana") original y en prosa letra de Gabriel Merino y Enrique López Marín, música de los maestros Caballero y Hermoso
A fuerza de arrastrarse	José Echegaray	Cervantes Virtual
Alfredo; Drama trágico en cinco actos	Joaquín-Francisco Pacheco	Cervantes Virtual
A Madrid me vuelvo	Manuel Bretón de los Herreros	Cervantes Virtual
Los Amantes de Teruel : drama en cinco actos en prosa y en verso	Juan Eugenio Hartzenbusch	Cervantes Virtual
Amor de padre	Francisco Martínez de la Rosa ; edición y estudio preliminar de Carlos Seco Serrano	Cervantes Virtual
Amor venga sus agravios	José de Espronceda	Cervantes Virtual
Ángela	Manuel	Cervantes Virtual

	Tamayo y Baus	
Epistolario de Gabriel y Galán	seleccionado por Mariano de Santiago Cividanes	Cervantes Virtual
Al primer vuelo	Pereda, José María de (1833-1906).	Edición digital a partir de *Obras completas. Tomo XVI,* Madrid, Viuda é Hijos de Manuel Tello, 1896.
El comendador Mendoza	Valera, Juan (1824-1905)	Edición digital basada en la de Madrid, Librería Enrique Prieto, [1906]
Hombre de mundo, Don Fernando el de Antequera, La muerte de César, La crítica de El Sí de las Niñas, & Fantasía dramática para el aniversario de Lope de Vega	Ventura de la Vega	Obras escogidas de Ventura de la Vega de la Real Academia Española. Tomo Primero
Las batuecas, comedia de magia en siete cuadros en verso y prosa h	de D. Juan Eugenio Hartzenbusc	Cervantes Virtual
La bola de nieve	Tamayo y Baus, Manuel (1829-1898)	Publicación: Alicante : Biblioteca Virtual Miguel de Cervantes, 1999 Nota: Edición digital a partir de Obras Completas, Madrid, Fax, 1947, pp. 618-719.
El alcázar de Sevilla	Blanco White, José María (1775-1841)	Alicante : Biblioteca Virtual Miguel de Cervantes, 1999 Nota: Edición digital a partir de la edición de Vicente Llorens, en José Mª Blanco White, Antología de obras en español, Barcelona, Labor, 1971, pp. 295-310.
Cosa cumplida _ solo en la otra vida : diálogos entre la juventud y la edad madura	Caballero, Fernán (1796-1877)	Alicante : Biblioteca Virtual Miguel de Cervantes, 1999 Nota: Edición digital basada en la edición de Madrid, Librería de Miguel Guijarro, 1881.
Esbozos y rasguños	José María de Pereda	Edición digital a partir de la de OO.CC., Madrid, Impta. de Manuel Tello, 1888, t. II y cotejada con la edición crítica de Salvador García Castañeda (OO.CC., Santander, Tantín, 1989, t. II, pp. 143-399)

El amigo Manso (Caps I-XXX)	Pérez Galdós, Benito (1843-1920(1882)).	Biblioteca virtual Miguel de Cervantes. Edición digital basada en la edición de Madrid, Administración de La Guirnalda y Episodios Nacionales, [1882]. Ejemp. de la Biblioteca Nacional (España)
El caballero de las botas azules	Castro, Rosalía de (1837-1885(1867))	Biblioteca virtual Miguel de Cervantes. Edición digital basada en la de Lugo, Imp. de Soto Freire, 1867.
El buey suelto	José María de Pereda (1884)	biblioteca virtual Miguel de Cervantes. Edición digital a partir de Obras completas. Tomo II, Madrid, Imprenta de Manuel Tello, 1884.
El cisne de Vilamorta	Emilia Pardo Bazán (1884)	Biblioteca virtual Miguel de Cervantes. Edición digital basada en la 1ª ed. de Madrid, Ricardo Fe, 1885
El copo de nieve (Caps I-V)	Ángela Grassi (1876)	Biblioteca virtual Miguel de Cervantes. Edición digital a partir de la edición de Madrid, Tipografía de G. Estrada, 1876.