

A usage-based approach to Spanish verbs of ‘becoming’

Joan Bybee and David Eddington

University of New Mexico and Brigham Young University

Address: Joan Bybee

MSC03-2130, LINGUISTICS

1 UNIV OF NEW MEXICO

ALBUQUERQUE, NM 87131

E-mail: [jbybee at unm](mailto:jbybee@unm.edu) edu

Telephone: 505-277-3827

Address: David Eddington

4041 JFSB

BRIGHAM YOUNG UNIVERSITY

PROVO, UT 84602

Email: eddington at byu edu

Telephone: (801) 422-7452

A usage-based approach to Spanish verbs of ‘becoming’

Abstract

A usage-based analysis of four constructions in Spanish, each with a different verb meaning ‘become’ used with an animate subject and an adjective, provides evidence for exemplar representations of constructions, with analogy to these representations accounting for productive use. We analyze 423 tokens from spoken and written corpora, which we take to represent a subset of a speaker’s experience with these constructions. The analysis, based on token frequency and semantic similarity leads to the organization of tokens with two of the verbs into dense clusters of semantically related adjectives centered on a high-frequency exemplar. The other two verbs are used with more diverse sets of adjectives. The initial analysis is supplemented with an experiment in which speakers were asked to rate the semantic similarity of pairs of adjectives. When subjected to multidimensional scaling, the results of the experiment support the initial analysis. We argue that novel instances of verb + adjective sequences are based on analogies to previous experience and not on rules that refer to abstract features. In a second experiment, speakers judged the acceptability of sentences taken from the corpora; the results showed that high-frequency expressions and expressions semantically similar to the high frequency ones lead to an expression being judged more acceptable. Overall the results support exemplar representations, which are heavily based on usage experience.*

* The authors are grateful to Rena Torres Cacoullos, Sandra Thompson, and the editors and reviewers of *Language* for comments and suggestions that led to improvements in the content and form of the paper. We are indebted to Halvor Clegg and the Humanities Research Center at Brigham Young University for making available to us the compilation of Spanish-language novels for use in the corpus study. We express our gratitude to José Antonio Mompeán for graciously volunteering to administer the questionnaires at the Universidad de Murcia. We also thank Ana Aurora Medina Murillo for her assistance with the semantic categorization, and Vsevolod Kapatsinki and Clayton Beckner for their help with editing and formatting.

1. Introduction. In this study we consider four commonly occurring Spanish verbs that express a change of state when used with adjectives and an animate subject. The verbs are all reflexive in form, and are given here with the glosses of the related non-reflexive form: *ponerse* ‘to put (reflexive)’, *volverse* ‘to turn (reflexive)’, *quedarse* ‘to remain (reflexive)’, and *hacerse* ‘to make (reflexive)’. Despite their etymological differences, all four verbs are now used with adjectives with a sense of becoming, for instance *ponerse nervioso* means ‘to get nervous’ and *quedarse sorprendido* means ‘to be(come) surprised’. In addition, there are three other verbal expressions for becoming that are used primarily with nominal complements and are not studied here. They are *llegar a ser* literally ‘to arrive at being’, *convertirse* ‘to convert (reflexive)’ and *transformarse* ‘to transform (reflexive)’. Given some degree of synonymy, the question naturally arises as to how a Spanish speaker chooses a particular verb for a particular situation. As might be expected, this issue has been a subject of some debate. Several researchers (Crespo 1949, Coste and Redondo 1965, Fente 1970, Eberenz 1985) have proposed criteria for distinguishing among these change-of-state verbs based on the type of change they describe. In particular, certain verbs are thought to express changes that are unexpected or unusual while others convey normal or expected changes. Other proposed factors have to do with whether the change involves an essential or nonessential property, or whether the change results in a temporary or permanent state. However, the particular features proposed, such as the degree of expectedness, essentiality, and permanence, are not easily defined.

Eddington (1999) applied four of the more concrete factors suggested in the literature to 1283 tokens of change-of-state verbs that occurred in a large natural corpus. The factors were: (1) whether the verb’s predicate is nominal or adjectival; (2) whether the change requires more than 24 hours to take place; (3) whether the change occurs with the active participation of the animate experiencer; (4) whether the noun or adjective of the predicate would be expressed with the

copulative *ser* or *estar*. A number of tendencies were found. For example, *llegar a ser* is generally used with changes that require more than 24 hours to complete. *Ponerse* and *quedarse* are followed almost exclusively by adjectival complements, while *convertirse* and *transformarse* take nominal complements. In spite of these tendencies, the data showed that the semantic features did not successfully determine the occurrence of the verbs, and moreover that occurrences of these seven verbs are not necessarily mutually exclusive; more than one verb may be appropriate in a given context.

Overlap among the uses of the change-of-state verbs is apparent in another study as well. Eddington (2002) presented a questionnaire to native Spanish speakers whose task was to indicate which change-of-state verb they would prefer in a sentence. One questionnaire item was the following (where the complement *alcohólicos* could be either a noun or an adjective):

(1) Muchas personas que beben van aumentando la cantidad que beben cada día, y sin darse cuenta: ('Many people that drink increase the quantity that they consume each day and without realizing it:')

___ se quedan alcohólicos. ('they become alcoholics')

___ se ponen alcohólicos.

___ se convierten en alcohólicos.

___ se hacen alcohólicos.

___ se transforman en alcohólicos.

___ se vuelven alcohólicos.

___ llegan a ser alcohólicos.

Sixteen different combinations of the four factors discussed above were incorporated into Eddington's test items, and participants marked all responses that they deemed appropriate. Responses to most questionnaire items centered on two of the seven possibilities. However, responses were quite varied, and an average of 1.5 different responses were marked per test item. This indicates that rather than there being complementary distribution between the uses of each verb, there is a great deal of overlap in their productive application to new situations.

These studies demonstrate the failure of categorical (all-or-none) features to predict the use of change-of-state verbs in Spanish. While some strong tendencies exist that allow one to differentiate certain uses of the change-of-state verbs, the boundaries of the categories of complements to these verbs are fuzzy. These results suggest that attempts to find general categorical features that characterize all uses of each verb and distinguish each verb from the others should be given up in favor of an approach that identifies some tokens of use as more central and others as more marginal. This approach would allow for the existence of some overlap at the boundaries of categories, but at the same time account for the distributional data that shows that some combinations are highly conventionalized and regular across speakers. Our study is couched within a usage-based exemplar framework in which it is claimed that the speaker's choice of a particular verb + complement combination is based on his/her accumulated experience with the use of these constructions.

Because our analysis does not attempt to find general features that cover all instances of use of a particular verb, but rather studies the specific relations found among tokens of usage, we have restricted our study to the four verbs that occur most frequently with adjectives—*quedarse*, *ponerse*, *volverse*, and *hacerse*—and we have only studied the examples in which these verbs occur with adjectives or prepositional phrases and an animate subject.¹ Since the vast majority of the tokens considered are adjectives and prepositional phrases constitute a small minority, hereafter

we refer to the constructions as consisting of a verb + adjective combination. The study is based on a large spoken and written corpus which we describe in detail below.

1.1. A usage-based exemplar model. A usage-based exemplar model categorizes the actual tokens of linguistic experience. As words, phrases, or constructions are used they are mapped by speakers and hearers onto identical existing representations if they are present, strengthening them. Such representations are called ‘exemplars’ and the term ‘token’ is reserved for the actual instances found in usage. For the purposes of this mapping, identity involves both semantic and phonological features.² If no identical exemplars are available the token is evaluated for its degree of similarity to other exemplars and represented as (metaphorically) close to or distant from existing exemplars. In this way, clusters of related items are built up. Such models have been applied to general categorization in the psychology literature (Medin and Shaffer 1978, Nosofsky 1988; see Chandler 2002 for an overview), to phonetic representations (Johnson 1997, Bybee 2001, Pierrehumbert 2001) and to constructions (Bybee 2006). In the present paper we apply such a model to the categorization of the adjectives and prepositional phrases that have been found in natural corpora occurring in constructions with four different Spanish verbs. Our analysis, then, is similar to that of Israel (1996), who studied the diachronic development of verbs that occur in the English *way* construction (*They dug their way out of prison*) and argued that novel instances of the construction were formed on analogy with existing instances.

As pointed out by Langacker (1987) for cognitive grammar, and Bybee (2001) and Pierrehumbert (2001, 2002) for phonology, the cognitive representation of language requires categorization of different levels of abstraction. At the phonetic level, a particular verbal form such as *me pongo* (1s Present Indicative of *ponerse*) might be represented with a set of phonetic exemplars. At a different, and perhaps higher or more abstract, level, all the inflected forms of the

verb *ponerse* are grouped together. At a still higher level, all the different exemplars of the construction *ponerse* + adjective are grouped together. The two-word combination that makes up one exemplar of the construction is taken to be represented in memory even though each of its two parts may also map onto the individual exemplars for *ponerse* and *nervioso* in the manner proposed for morphology in Bybee (1985).

In the forms examined here, parts of a token of a construction might be identical to existing representations and parts of it different, as in Spanish *ponerse nervioso* ‘to get nervous’ and *ponerse furioso* ‘to get angry’. In such cases, the identical parts are mapped onto the same representation, strengthening it, while the non-identical parts are grouped together, forming a category with other adjectives that occur with *ponerse*. It is the resulting category of adjectives that is the focus of our analysis. We find that an instance-based or exemplar model of categorization which takes into account both token and type frequency in the corpus works well for explaining the range of adjective types found in the corpus for each construction.

In its simplest form, exemplar-based categorization takes place by comparing some incoming item or probe to similar exemplars stored in memory and then assigning this probe to one of the stored exemplars, strengthening it. This type of categorization can theoretically proceed with no permanently stored categories; rather an analogical set of similar exemplars is formed ad hoc for each probe (Skousen 1989, Chandler 2002). However, it is not incompatible with such a model to assume, as we do here, that a frequently calculated analogical set is stored in long-term memory. Because linguistic categories are so frequently used, we find it plausible to assume that they exist in long-term memory.

In contrast, a prototype model creates categories by extracting features from instances, but does not store the exemplars upon which the categorization is based (Taylor 1995: 59ff). When compared, exemplar models generally outperform models based on prototypes (Chandler 2002,

Murphy 2002). In our example of the adjectives that follow *ponerse*, the exemplar model would simply list the adjectives, organized by similarity, as in 2.

(2) ponerse	$\left[\begin{array}{c} \text{nervioso} \\ \text{lívido} \\ \text{pálido} \end{array} \right]$	'nervous'
		'pale, purple'
		'pale, pallid'

A prototype model would not list the actual adjectives, but would rather extract features from them, as in 3.

(3) ponerse	$\left[\begin{array}{c} \text{adjective} \\ \text{features X, Y, Z...} \end{array} \right]$
-------------	----------------------------------------------------------------------------------------------

The data we examined are naturally occurring instances in which speakers choose to use a certain adjective and a particular change-of-state verb to accompany it, so they must categorize the adjective in order to choose a verb. The examples we extracted from the corpus show that in most cases similarity to previous uses is an important criterion, but also that groups of similar adjectives cluster around high-frequency adjectives, suggesting that the token frequency of certain verb + adjective combinations is also an important factor.³

The data point to an exemplar model which registers tokens of experience. In some categories, exemplars with higher token frequency appear to function as central members. In these cases, high-frequency exemplars could be thought of as prototypical since they exhibit most of the features shared by the other members. To account for the data, individual exemplars must be stored in memory in order for their frequency in experience to be registered. At the same time, the

categories of adjectives seem to be conventionalized and stable in the language rather than composed ad hoc for each use of the verb + adjective combination. The best fit with the data is a model in which exemplars are stored and accrue strength on the basis of frequency while also being organized into categories around a central exemplar.

Both exemplar and prototype categorization share the following properties: there is no set of necessary and sufficient conditions for defining the category, members may be more central or more marginal, and the boundaries of the category are not discrete. In addition, both exemplar and prototype categories have a family resemblance structure (Wittgenstein 1953, Rosch and Mervis 1975, Lakoff 1987); category members share features (most commonly) with the central member or with some other member, but it is possible that marginal members may not share any features with one another. A linguistic category that has this structure is the class of irregular past tense verbs in English which has *strung* as its central member. A chain of family resemblances can be identified in the final consonants of the verbs *won* – *strung* – *struck* / *dug*. *Struck* and *dug* have nothing in common with *won*, but *won*, *struck*, and *dug* all share a feature with *strung* (Bybee and Moder 1983).

A major argument for exemplar theory over prototype theory is that exemplar representations allow for redundant or marginal features to serve as the model for novel expansions of the category (Chandler 2002). Since all features of an exemplar are represented, there is no need to predict in advance which features are chosen for the basis of expansion. This allows for changes in the category. For instance, the verb class mentioned just above, Class III, was defined in Old English as having a consonant cluster at the end of the stem. A large majority of verbs in this class had a stem-final geminate nasal or cluster of a nasal and obstruent, and the short vowels [] in the base, [] in the Preterit, and [] in the Past Participle, e.g. *swimman* ‘to swim’, *winnan* ‘to win’,

ge-limpan ‘to happen’, *bindan* ‘to bind’, *wringan* ‘to wring’ (Sweet 1965 [1882]). When vowel lengthening occurred before homorganic coronal clusters, as in *bindan*, the group of verbs with such clusters no longer fit in the category. Now, while the characterization of the category as having verbs ending in nasals was still valid, a previously unused feature, that of having a final velar consonant, became important to the category presumably because verbs with the velar consonant now constituted the majority of types, despite not appearing in all members. Other examples are *drink*, *shrink*, and *spring*. Other verbs ending in velars and nasals joined the class, e.g. *fling*, *sling*, *sting*, and *string*, further strengthening the velar nasal center of the category (Jespersen 1942). The importance of velars as a defining feature of the central members became apparent when verbs with final non-nasal velars, such as *dig* and *strike*, joined the class.

The model we have chosen allows representations to change as language is used because every token of experience has an impact on cognitive representation. Although a model with abstract prototypes could be updated continuously, the fact that redundant features are sometimes chosen as the basis of extensions of the category, means that it may not always be possible to predict which features should be retained as characterizing the prototype and which should be discarded. For this reason, we opt for an exemplar theory, where redundant features are retained in representation, at least for a while.

1.2. Usage-based construction grammar. By ‘construction grammar’ we refer to several recent proposals of grammatical theories that take the basic unit of morpho-syntax to be constructions — ‘stored pairings of form and function including morphemes, words, idioms, partially lexically filled and general linguistic patterns’ (Goldberg 2003).⁴ The notion of storage is important in this conception of grammar. Specific constructions involving more than one word are considered to be stored in memory, so that grammar and lexicon are interwoven. In a sense, morphosyntactic

constructions are just big words with moveable parts (Bybee 1998, Culicover 1999). In fact, it is possible to model syntax in terms of analogy to a mental lexicon that stores past linguistic experience (Bod 1998). It is notable that most of the constructions identified in the construction grammar literature involve specific words or morphemes, as for instance in morphological constructions where specific inflectional morphemes must be mentioned, and in syntactic constructions such as the one exemplified in *He made his way through the crowd*, where the word *way* appears with a genitive pronoun referring to the subject.

Most construction-based theories recognize any linguistic patterning of words or morphemes as a construction if there is some aspect of form, function, or meaning that is not predictable from either the component parts of the pattern or from other constructions in the language (Fillmore et al. 1988, Goldberg 1995). However, in a usage-based approach to constructions, a string of words or morphemes that is used with some frequency would also be considered a construction even if its form and function or meaning are entirely predictable. This assumption is necessary to understand how constructions that do have unpredictable features arise: in order to assign a special pragmatic function to a string of words, or to register a phonological change, that string of words must constitute an item in cognitive storage, a representation to which changes in function or form can be attached (Bybee 2001, 2006). This assumption follows from the premise that in addition to being units of storage, constructions are neuromotor processing units (Boylard 1997, Bybee 2003). Their repetition by the language user helps to automate their production. All the items in Goldberg's definition—words, idioms, and both partially filled and fully general linguistic patterns—are highly practiced sequences that can be accessed for production and comprehension.

Constructions contain fixed units, that is, particular words or morphemes that characterize the construction and, in addition, they may also contain variables or open slots which take a class

of items. The meaning of the construction is determined by the component parts and by the contexts in which the construction has been used. In an exemplar representation, all instances of use of a construction in an individual's experience would contribute to the meaning of the construction. Thus in the case of the constructions to be considered here, *quedarse* + ADJ, *ponerse* + ADJ, *hacerse* + ADJ, and *volverse* + ADJ, we ask not so much what the verb means, but what the overall construction means. That meaning can be arrived at only by considering the instances of use of these constructions. Moreover, the meaning resides in the exemplar clusters created by the user's experience and may not be describable by a set of abstract features that belong specifically to the verbs or the adjectives. For this reason, our analysis does not necessarily result in a definition of the verbs in the construction, but rather a characterization of the uses of the constructions as whole units, taking the adjective categories as the starting point in the analysis. As Croft (2001) points out, the categories that define the open slots in constructions constitute the grammatical categories of the language. In some cases below, we find it appropriate to comment on the contribution of the verb's meaning to the construction, but for the present study we have found the adjective categories interesting in their own right and leave further specification of the verbs' meanings for future study.

1.3. The goals of the paper. The organization of the exemplar clusters derived from a user's experience with a construction is the focus of analysis in this paper. By exploring the items that appear in the ADJECTIVE position in these constructions, we produce a conceptual clustering that can predict subsequent uses of the constructions. The data we analyzed (see Section 2.1) are from several million words of naturally occurring spoken and written language and thus simulate a fraction of the experience that a Spanish speaker might have with the constructions to be analyzed here, though we cannot of course know what the precise experience of any given speaker is and we

recognize that individuals have different experience with language. Our analysis attempts to map the tokens and types found in the corpora onto a representation that might plausibly represent a language user's cognitive representation of these constructions. An important aspect of this analysis consists of judgments of similarity in meaning among the adjectives that appear in the constructions. The intuitions of a native speaker were elicited to organize the adjectives into categories based on meaning. This analysis is presented and discussed in sections 3-6, with a summary in section 7. Aspects of the resulting semantic analysis were tested in an experiment with a large group of native speakers. The results are presented in section 8.

The second issue addressed in this paper concerns the nature of judgments of acceptability predicted by an exemplar-based construction grammar. As we will see in the analysis in sections 3 through 6, certain instances of each construction were more frequent than others. *Ponerse nervioso* 'to get nervous' and *quedarse quieto* 'to calm down' occurred multiple times in the corpus and were assumed to be robustly represented in linguistic memory, while other exemplars, such as *ponerse revoltoso* 'become rebellious' and *quedarse orgulloso* 'become very proud' were rare. Given exemplar representation, we hypothesized that the high-frequency exemplars would be judged more acceptable by native speakers than the low-frequency ones. The results of this experiment, which are in accord with the predictions of exemplar representation, are reported in section 9.

2. The corpus study.

2.1. Description of the corpora.

1 Instances of the verbs *quedarse*, *ponerse*, *volverse*, and *hacerse* were identified in two corpora. The first is a corpus of 1.1 million words of spoken Spanish (Marcos Marín 1992) that consists of transcribed spoken speech from Spain. This includes a wide variety of types of oral speech (e.g. patient/doctor interactions, sports commentaries, informal

conversations, classroom lectures). Instances of change-of-state verbs were also gleaned from written sources. We consulted a compilation of 15 novels from various Spanish speaking countries or regions (Argentina, Chile, Colombia, Cuba, Guatemala, Mexico, Peru, Puerto Rico, Spain, and Venezuela). All of the novels were published in the last half of the 20th century. This particular corpus of approximately 990,000 words was chosen simply because it was readily available.

Instances of change-of-state verbs that were followed by adjectives and prepositional phrases, and whose subject was human, were extracted with the use of a computer program called Word Cruncher. The study was restricted to expressions with human subjects so that the number of types and tokens would be manageable and because a preliminary analysis suggested this restriction would yield more coherent categories. Each instance encountered was inspected in order to insure that it denoted the inception of a new state. *Quedarse*, for example, can indicate either inception or continuation of a state previously entered into: *Se quedó con hambre*, ‘He was still hungry’.

Cases in which inception was not apparent from the context were discarded for the purposes of the analysis. Similarly, for *hacerse*, cases in which the meaning was ‘act’ rather than ‘become’ were not analyzed.

2.2. The nature of the analysis. Our analysis follows closely the distribution of the tokens in the data and takes into account both the token frequency of certain exemplars of the constructions and the distribution of the construction over various types. As general background, Table 1 presents the type and token frequencies for the four ‘become’ verbs in the corpora analyzed. The analysis is based on the 423 tokens of use found in the corpora.

INSERT TABLE 1 ABOUT HERE

The presentation proceeds from the most frequently occurring constructions to the least frequent.

Two important facts about the data led us to the type of analysis we pursued. First, in many of the semantic domains covered by the adjectives that occurred in the corpora, we observed a certain pattern in the data: a particular combination of verb + adjective occurred quite frequently. These combinations constitute ‘prefabs’ or fixed expressions that are conventionalized, and in this case frequently used (Pawley and Syder 1983, Sinclair 1991, Erman and Warren 2000, Wray 2002). Here are some examples from the data:

(4)	quedarse solo	‘to end up alone’
	ponerse nervioso	‘to get nervous’
	quedarse tranquilo	‘to calm down’
	volverse loco	‘to go crazy’

Many of the other less-frequently used adjectives were synonymous or shared features with the adjectives that occurred in these conventionalized expressions, which led to an analysis in which these adjectives constituted the central members of categories that fanned out from these centers (Lakoff 1987). Instead of searching for features that characterize all the adjectives that occur with each verb, we organized the adjectives into multiple similarity groups. In the following sections, we present this analysis.

Second, very few adjectives occurred with more than one verb. Given the distribution mentioned above, this is to be expected. Since there were conventionalized verb + adjective combinations that occurred frequently and less frequent uses were clustered around these, the verbs were rarely used productively to suggest nuances of meaning in the process of becoming, for instance, creating contrast between types of becoming with the same adjective. As Table 2 shows,

of the forty-eight types that occurred more than once in the data (accounting for 332 tokens), only twelve of these occurred with different verbs (accounting for 67 tokens). Therefore, contrasts of the same adjective used with different verbs occurred with only one quarter of the lexical types.

INSERT TABLE 2 ABOUT HERE

2.3. Steps in the analysis. The adjective tokens that occurred with each verb were roughly organized into similarity groupings by the authors without making a distinction between tokens that occurred in the spoken versus the written corpus. To check and refine these groupings, a native speaker from Mexico, Ana Aurora Medina Murillo, was asked to make judgments concerning the semantic similarity of the adjectives. She was not informed as to the relative frequency of the adjectives, nor was she asked to use frequency as a criterion. She was given each adjective written on a post-it note and asked to position these on a blank sheet of paper, according to their semantic similarities. In some cases it was necessary for the consultant to examine the context of the verb + adjective combination in the corpus to determine its semantic affinities. The resulting groupings will be presented below in the figures in sections 3 through 6.

The resulting analysis was then tested experimentally in two different ways. In the first experiment we tested our similarity groupings of a subset of the corpus adjectives by asking 77 native speakers for similarity ratings which were then submitted to a multidimensional scaling analysis. The results of this experiment are presented in section 8. In the second experiment we examined the effect of token frequency on category membership by asking subjects for acceptability judgments on corpus sentences that contained high-frequency expressions, low-frequency expressions related semantically to high-frequency expressions, and low-frequency expressions not related to high-frequency expressions. The results of this experiment are presented

in section 9. In both cases, the results of the experiments present striking confirmation of the analysis that emerged from the corpus analysis.

We turn now to the analysis of the corpus materials.

3. *Quedarse* + ADJECTIVE. The analysis of the adjectives occurring with *quedarse*, the most frequently used of the verbs in the constructions studied here, resulted in five distinct categories, some of which have fairly complex internal organization. We begin with a rather simple category, the one centered on *quedarse solo*, in order to illustrate our method of analysis.

3.1. *Quedarse solo* ‘to become alone’. Table 3 presents the types and their token frequency in the spoken and written corpora for exemplars of *solo* ‘alone’ and related adjectives. The following are two representative examples from the database.

(5) 1Menelao **se queda solo**.

‘Menelao is left alone.’

(6) **1**Se **quedó soltera** por falta de padre o de hermano.

‘She became a spinster for lack of father or brother.’

The adjectives in Table 3 were grouped together by the authors and by the consultant. The most frequent adjective, *solo*, was positioned near the center and it is hypothesized to be the central member of the category. The last adjective listed means ‘paired with’ and appeared in a sports commentary in which two players appear to be faced off against each other. We include it in this

cluster as an opposite on the assumption that opposites share a number of features while having a negative value for one important feature.

INSERT TABLE 3 ABOUT HERE

Figure 1 supplies a graphic illustration of the cluster of adjectives in Table 3. The figures are meant to illustrate impressionistically the relations among the adjectives; they are not meant to represent a strict formalism nor are they intended to picture the one and only categorization possibility. *Solo* is placed in the center to indicate its hypothesized status as the central member; the darker shading and the larger font are intended to impressionistically indicate its higher frequency and central status. The opposite adjective *emparejado* is set off by a line with arrows on each end, a convention we have chosen to indicate that it has an opposite value in one feature and constitutes not a synonym or related equivalent, but nonetheless a related concept.

INSERT FIGURE 1 ABOUT HERE

3.2. *Quedarse quieto* ‘to become calm’. With these principles of analysis and representation laid out, we now proceed to view one of the most complex clusters of adjectives identified in the data and the one that encompasses more types and tokens than any other. This cluster is complex because there are five clusters that overlap and are related. Figure 2 shows the most frequent members of these five clusters and how they have been positioned by the native-speaker consultant. *Quieto* ‘calm, still’ has been chosen as the central member as it shares features with *callado* ‘quiet’, *inmóvil* ‘still, motionless’, and *tranquilo* ‘calm, peaceful’. Not coincidentally, we would claim, *quieto* has the highest token frequency in this construction of all the adjectives,

although it occurred only in the written corpus (see Table 4). *Dormido* ‘asleep’ prototypically shares features ‘calm’ and ‘still’ with *quieto*; however, our consultant felt it should be positioned nearby in this metaphorical space, but that it did not overlap with the other adjectives. Here we see a chain of family resemblances. Since *quieto* indicates stillness, quiet and inferentially peacefulness, each of the other adjectives shares a feature with it: *callado* (absence of sound), *inmóvil* (absence of motion), and *tranquilo* (peacefulness). *Callado* and *inmóvil* each have less in common with each other than each has with *quieto*. *Tranquilo* introduces a different feature, one that suggests some contentment; this in turn links with *a gusto* ‘pleased’ which has very little in common with *callado* or *inmóvil*.

INSERT FIGURE 2 ABOUT HERE

The frequencies of these central adjectives are given in Table 4. Each of these adjectives is the most frequent member of a cluster of adjectives with similar meaning, as shown in Figure 3 where all the corpus adjectives that were grouped in this set of clusters are pictured; Tables 5a through 5g list all the adjectives in each subcluster. Note that several adjectives occur in the overlapping space of two clusters; thus their listing in the tables was somewhat arbitrarily chosen. Although almost all the adjectives are related to some degree to *quieto*, still for convenience, only the adjectives that are related to *quieto* more than to any other central member are listed with *quieto* in the tables.

INSERT TABLE 4 ABOUT HERE

As can be observed in the tables, the choice of the most frequent member of a cluster is quite obvious, as the frequency differences are typically quite robust. It is indeed this fact that led to the

nature of the current analysis. The lowest frequency exemplar listed in Table 4 is *a gusto* which occurred only four times. This adjective is distant enough in meaning from *tranquilo* to suggest to our native-speaker consultant a new cluster of which it is the central member.

INSERT FIGURE 3 ABOUT HERE

Figure 3 shows how the non-central members form clusters around the central members.

The following are some selected examples of central and non-central members:

- (7) Al verme **se quedaron quietos**.
‘When they saw me they calmed down.’
- (8) **1Se había quedado dormida** sin apagar la vela.
‘She had fallen asleep without blowing out the candle.’
- (9) **1No les había pedido que se quedaran callados**.
‘He hadn’t asked them to quiet down.’
- (10) **1Es una prueba que necesitan hacer ellos para quedarse a gusto de** que el . . .
‘It’s a test they need to do in order to feel pleased that the . . .’
- (1) **Pareces un perrito manso cuando te quedas pensativo**--dijo ella.
‘You look like a gentle dog when you get pensive, she said.’

(12) Lo que ahora va a ocurrir es ese mal momento en que, no atreviéndose ninguno de los dos a iniciar un ademán cualquiera, **nos quedamos trabados**, a punto de decir alguna tontería.

‘What's going to happen now in that bad instant, in which since neither of us dares to make any sort of gesture, we become stuck on the verge of saying something stupid.’

INSERT TABLE 5a – 5g ABOUT HERE

The figures and the tables represent category members of four types: (i) central members, which have high token frequency and which are closely related to other members; (ii) adjectives that have a part of their meaning very similar to a central member (e.g. in one sense *satisfecho* ‘satisfied’ is very similar in meaning to *tranquilo* ‘peaceful’ in that the person whose need goes unsatisfied is not at peace.); (iii) metaphorical extensions of the concept expressed in the central member (e.g. *de piedra* ‘of stone’ as related to *inmóvil* ‘immobile’); and (iv) opposites (*despierto* ‘awake’ as an opposite to *dormido* ‘asleep’). The examination of other verbs reveals that there are semantically isolated uses of adjectives as well, though these constitute a small minority of tokens.

3.3. *Quedarse sorprendido* ‘to be surprised’. The arbitrariness of the conventionalized combinations found in the central members is especially clear in the next major cluster of adjectives used with *quedarse*, which centers on *quedarse sorprendido* ‘to be surprised’. While *solo* ‘alone’ and *quieto* ‘calm’ might seem to have something in common, i.e. an absence of movement, agitation, or company, *quedarse sorprendido* does not share these features, nor is it strictly an opposite, as it focuses on an abrupt reaction to the presentation of new information, a state that wears off rather quickly. Therefore, it appears that the various categories of adjectives

used with *quedarse* need not be related to one another. However, the internal structure of this grouping has the same properties of the groupings just examined. See Figure 4 and Table 6.

INSERT FIGURE 4 ABOUT HERE

- (13) 1Al voltearlo con la puntera de la bota para alumbrarle la cara, el capitán **se quedó perplejo.** //

‘Upon turning him over with the tip of his boot in order to shine some light on his face the captain became perplexed.’

- (14) 1Ahora creo que los pivots españoles **se han quedado un tanto sorprendidos.**

‘Now I think the Spanish centers have become a little surprised.’

- (15) 1Pues, esta persona se metía dentro de la cueva y **se quedaba alucinado.**

‘So, this person was getting inside the cave and was amazed.’

INSERT TABLE 6 ABOUT HERE

Figure 4 represents *sorprendido* ‘surprised’ as the central member of a large cluster. *Frío* is used as in example 16 and indicates fear as well as surprise.

- (16) Lo he tenido en mis manos hace poco y **me quedé eh- frío** al ver la precision.

‘I had it in my hands a while ago and it left me, uh, cold to see the precision.’

The other examples in this category appear to be quite closely related, all indicating surprise, amazement, astonishment, and fear. The one use of *loco* ‘crazy’ here, which is uncharacteristic since *loco* is conventionally used with *volverse*, comes in a context in which the discussion is about how the prices of retail items vary from one place to another. The speaker then says:

(17) Bueno y las Ray Ban ayer **me quedé loca**.

‘Well, the (prices of the) Ray Bans yesterday drove me crazy.’

This seems to indicate that the wide variation in the prices of the Ray Ban (sunglasses) both astonished her and drove her crazy.

3.4. *Quedarse triste* ‘to become sad’. *Quedarse triste* occurs four times in the written corpus and once in the spoken corpus. The spoken corpus contains three other examples that express concepts that appear related to *triste* as shown in Figure 5 and Table 7

INSERT FIGURE 5 ABOUT HERE

INSERT TABLE 7 ABOUT HERE

The following is an example:

(18) En el fondo **se quedaban un poco tristes** pensando en posibilidades malogradas por el carácter argentino y el paso implacable del tiempo.

‘In their hearts they got a bit sad thinking about possibilities unachieved by the Argentine character and the unstoppable passing of time.’

Triste also appears as the center of a cluster of adjectives occurring with *ponerse* as we discuss in section 4.2. It is the only adjective that is used with more than one verb enough to merit it as the center of two clusters (both of which are rather small in terms of both type and token frequency). We will compare the two clusters in section 4.2.

3.5. *Quedarse* + physical state. The conventionalized way of saying ‘become pregnant’ for humans in Spanish is *quedarse embarazada*, which occurred four times in the spoken corpus. For animals the expression is *quedarse preñada*, which occurred three times in the spoken corpus and once in the written corpus. Other tokens that the consultant grouped with these terms also described physical states, but none were considered to form a cluster with *embarazada*. They are listed in Table 8. We have not provided a figure in this case because the adjectives are distributed in separate, non-overlapping areas of conceptual space.

INSERT TABLE 8 ABOUT HERE

Examples from this cluster are the following:

(19) Celeste **se ha quedado ciega**.

‘Celeste has become blind.’

(20) 11Fue muy duro todo, el **quedarme embarazada** tan rápidamente, y tener otro hijo tan seguido.

‘Everything was really tough, getting pregnant so fast and having another child so soon.’

These adjectives are the only ones used with *quedarse* that might be considered isolated, but even these have some relation to *embarazada* and to one another through their reference to physical states. In other words, we find that the productivity of *quedarse*, which is the most frequent verb meaning ‘become’ in the database, is highly dependent upon relations with adjectives with similar meanings. Less frequent uses of *quedarse* appear to be formed on analogy to existing higher frequency conventionalized uses.

4. *Ponerse*. In the data we identified four sets of adjectives used with *ponerse*, and as with *quedarse* some of the clusters showed highly structured relations, while others appear to be looser in their organization. In addition, there were some uses that were not expanded upon in the corpus (e.g. *ponerse ciego* ‘to become blind’ and *ponerse desnudo* ‘to get naked’), and a few more or less isolated examples.

4.1. *Ponerse nervioso* ‘to become nervous’. One of the main clusters involving *ponerse* centers on the highly conventionalized phrase *ponerse nervioso*, which occurred frequently in the corpora examined. As Figure 6 shows, the native speaker consultant suggested three groupings in this cluster, as well as indicating that *nervioso* and *histérico* were related.

INSERT FIGURE 6 ABOUT HERE

The state of being nervous includes varying degrees of agitation that can be expressed in various ways, such as by becoming red in the face, or by turning pale or becoming rigid or stiff.

Two groups of adjectives (found primarily in the literary corpus) describe the internal state of agitation with reference to these physical changes.

The relation between *nervioso* and *histérico* seems to involve the degree of agitation, with *histérico* describing an extreme state of agitation, though its source is not included in the meaning. Another set of adjectives expresses agitation but is explicit about the source being anger: *furioso*, *agresivo*, and *bravo*. *Bravo* means ‘angry’ as example 21 shows:

(21) Nada más que una broma, capitán, no **se nos ponga**⁷ **bravo** ahora que estamos más tranquilos.

‘Just a joke captain, don’t go getting mad now that we’re more calm.’

Tables 9a-9c show the frequencies of occurrence of the types in this cluster.

INSERT TABLE 9a-9c ABOUT HERE

The following are examples from each subgrouping:

(22) ¡Qué **pálida se ha puesto**!⁸

‘She’s really become pale!’

(23) 1De repente **se pone** furiosa porque he mirado dos veces a una persona.

‘All of a sudden she gets furious because I looked at someone twice.’

4.2. *Ponerse serio* ‘to become serious’. Another smaller grouping of adjectives that occurred with *ponerse* centers on *serio* ‘serious’ and *triste* ‘sad’ and shows an organization in which the adjectives appear highly related semantically to one another. (See Table 10 and Figure 7).

INSERT TABLE 10 ABOUT HERE

INSERT FIGURE 7 ABOUT HERE

This grouping contrasts with the small grouping centered on *triste* with *quedarse* in that those adjectives (*nostálgico* ‘nostalgic’, *hecho polvo* ‘pulverized, depressed’, and *como un estropajo* ‘like a scourer, battered’) are more focused on subjective and local personal emotional states, which the subject has entered as the result of external circumstances as in the following examples from the spoken corpus. Perhaps there is a relation here with the grouping *quedarse sorprendido* ‘become surprised’ in that external circumstances impinge upon the subject, creating the reaction.

(24) Ahora me hace recordar, **me quedo un poco nostálgico**.

‘Now it makes me remember and I get a little nostalgic.’

(25) Yo **me he quedado muy triste** porque no podemos regular las consumiciones para el concierto.

‘I’ve gotten really sad because we can’t regulate the expenditures for the concert.’

In contrast, the adjectives in Table 10 used with *ponerse* suggest involvement of the subject in entering the state.

(26) Y el Mochuelo **se puso**⁹ **más triste** todavía, pensando que cuatro semanas después él se iría a la ciudad . . .

‘And Mochuelo got even sadder still thinking that four weeks later he’d go to the city ...’

(27) Y, volviéndose a mí, **se puso triste**:—Lo que me importa es que van a separarnos y no te podré ver nunca más.

‘And turning toward me she became sad. “What matters to me is that they are going to separate us and I won’t ever be able to see you again”.’

(28) No hemos hablado de cosas así—tremebundas porque ninguna de las dos estábamos como para **ponernos trágicas**.

‘We haven’t talked about things like that, terrifying things, because neither of us was in the mood to get tragic.’

These adjectives with *ponerse* seem to designate a manner of behaving as much as an emotion which a person could be subject to upon thinking of certain matters.

This difference between *quedarse* and *ponerse* with these adjectives reflects their original lexical meaning, much as lexical meaning can be retained in grammaticization (Bybee and Pagliuca 1987, Hopper 1991). *Quedar* means ‘to remain or stay, to be left or left over’, a fairly passive sense as others have noted (Crespo 1949, Coste and Redondo 1965, Fente 1970, Eberenz 1985). *Poner* is more active, meaning ‘put or place’. *Ponerse* would literally mean ‘to place oneself’, a meaning that indicates a more active role for the subject. While the feature of subject agentivity evident in these examples has been suggested for this distinction before, it is important to note that it is neither a necessary nor a sufficient condition for the use of *ponerse*, since there are

many instances where no agentivity is involved: *ponerse nervioso* ‘nervous’, *pálido* ‘pale’, *rojo* ‘red’, *mal* ‘sick’, *ciego* ‘blind’. Moreover, some examples with *quedarse* seem quite parallel to examples with *ponerse*. For instance, 29 seems similar to 26 in degree of agentivity:

- (29) En el fondo **se quedaban un poco tristes** pensando en posibilidades malogradas por el carácter argentino y el paso implacable del tiempo.
 ‘In their hearts they got a bit sad thinking about possibilities unachieved by the Argentine character and the unstoppable passing of time.’

It seems that these verbal constructions, much like grammaticizing constructions, have uses in which much of their original meaning, in this case agentivity, is retained as well as uses in which that meaning is no longer discernible.

Serio occurred once with *quedarse*, but in that case it occurred in a conjoined phrase with *quieto* ‘calm, still’. Our consultant placed it with the adjectives in Figure 2 positioning it close to *pensativo* ‘pensive’.

4.3. *Ponerse pesado* ‘become annoying’. A grouping of adjectives describing decidedly negative personal behavior has as its most frequent type *pesado* ‘annoying’. The adjectives with their token frequencies are given in Table 11. Several groupings for these adjectives could be proposed, some overlapping slightly and some not. The exact structuring of this category is less apparent than for some of the others. We leave this issue for future research.

INSERT TABLE 11 ABOUT HERE

The relation of this group with the others used with *ponerse* seems to be somewhat arbitrary. One might suggest in this case that some agentivity is assigned to the subject, presuming that a person can control how annoying s/he becomes.

4.4. *Ponerse bonito* ‘to become pretty’. The conventionalized way of describing states that one more or less grows into, such as girls becoming pretty or boys handsome (*guapo*) uses *ponerse*. In the written corpus this usage is extended to the opposite, *feo* ‘ugly’ as well as to *gordo* ‘fat’ and *viejo* ‘old’. The organization of these adjectives is less obvious than in some of the other cases so no specific grouping is proposed here. Note that no volitionality is implied in these cases. The adjectives in this group each occurred only once in the corpora.

4.5. *Ponerse mal*, ‘to become sick,’ *ponerse contento* ‘to become happy’. The conventionalized way of saying that a person gets sick is to use *ponerse mal* ‘to get sick’ (literally ‘bad’), which occurred six times in the corpora. A synonymous expression, *ponerse enfermo* ‘get sick’, occurred four times in the written corpus. *Ponerse terrible* ‘to become terribly [sick]’ was also grouped with these expressions.

Ponerse contento appears to be the conventional way to express becoming happy. It occurred four times in the written corpus.

4.6. Non-conventionalized isolated uses and their meanings. Five examples with *ponerse* did not fit into the groups established so far, nor do they exhibit semantic similarity among themselves.

Nevertheless, our consultant grouped three of them together saying that the implication with these adjectives was that the subject of the verb was said to start acting a certain way. The adjectives are *culto* ‘cultured’, *maternal* ‘maternal’, and *putona* ‘sleazy’. In these more productive uses of *ponerse* the sense of the subject’s volitionality is clear.

- (30) 1Desgonzado y temblando y **te pusiste maternal**, pegada a mí, acariciándome la frente.
 ‘Out of control and trembling you acted maternal, sitting right next to me and caressing my forehead.’

Two other expressions occurred in the corpus. The first is *ponerse ciego* which means ‘to go blind’. It occurred once in its literal sense in the spoken corpus and twice in the figurative sense *ponerse ciegos de dinero* ‘to become blinded by money; filthy rich’. The second expression, *ponerse desnudo*, ‘to get naked’ occurred twice in the spoken database and seems to be an isolated use.

4.7. Summary of *ponerse*. Out of 42 different adjectives used with *ponerse* only five are not formulaic expressions or included in a semantic category (most of which are themselves centered around a formulaic expression). In these cases, and a few others, *ponerse* indicates that the subject is acting a certain way. As pointed out above, this sense is not apparent in the majority of uses. In the majority of uses the choice of *ponerse* is conventionalized and the meaning it expresses is simply one of becoming.

5. *Volverse*. The distribution of adjectives with *volverse* and *hacerse* (to be discussed in the next section) is quite different from the distribution seen with *quedarse* and *ponerse*. *Volverse* and *hacerse* have a much lower type and token frequency when used with adjectives and they have fewer formulaic uses with animate subjects.¹⁰

Volverse appears to have only one formulaic use, though it is a frequent one: *volverse loco* is the conventionalized way of saying ‘to go crazy’. This phrase appeared six times out of only seven tokens in the spoken corpus and ten times out of 24 tokens in the written corpus. The other adjectives used with *volverse* each occurred only once and they were largely unrelated in meaning to *loco*. (See Table 12.) Outside of its use with *loco*, *volverse* seems to constitute a generalized possibility for forming ‘become’ expressions, especially in written language.

INSERT TABLE 12 ABOUT HERE

6. *Hacerse*. *Hacerse* is more commonly used with inanimate subjects, where it enters into formulaic expressions such as *se hace tarde* ‘it’s getting late’, and with noun complements, as in *hacerse amigos* ‘to become friends’. With animate subjects and adjective complements it demonstrates a rather broad productivity, forming no strong clusters. Table 13 shows expressions with *hacerse* that occurred in the corpora.

INSERT TABLE 13 ABOUT HERE

7. Summary of findings with the four verbs. The natural data examined here show that the use of ‘become’ verbs with adjectives in Spanish is heavily dominated by formulaic uses or fixed phrases. Numbers can be calculated for this trend by counting the number of single occurrences compared

to the number of tokens that represented a type that occurred more than once. This count will give us an approximate number for fixed expressions vs. novel uses. It is only approximate because it could be that a fixed expression happened to occur only once in the corpus or a novel expression unique to a single author might have occurred more than once. Expressions with *quedarse* used only once in the data constituted only 21% of the expressions with this verb. Single occurrence expressions with *ponerse* constituted 24%; with *hacerse*, 58%; and with *volverse*, 34%. Overall 25% of the tokens were novel instances and 75% belonged to types that occurred more than once. Therefore, it appears that the manner of expressing becoming in Spanish is highly conventionalized.

In addition, a large majority of the tokens are semantically related to these central exemplars, suggesting that they are created via analogy based on semantic properties. Most commonly these analogies are based on synonymy or near synonymy (*callado* ‘quiet’ is closely related to *quieto* ‘quiet, still’) but they may also be metaphorical, as when *de piedra* ‘(made) of stone’ is used to mean *inmóvil* ‘immobile’. Truly isolated uses—those that did not bear a semantic resemblance to any other exemplars of the construction used—constituted only 7% of the tokens in the corpora. These uses included the physical state uses of *quedarse*, the five instances of *ponerse* discussed in 4.6, uses of *volverse* not related to *volverse loco* and uses of *hacerse* that seemed unrelated to other uses.

In sum, the overall picture presented by the data is that there are certain well-established pairings of verb and adjective and that these are referred to by speakers in the choice of a ‘become’-verb for the expression of a particular state described by an adjective. As mentioned earlier, many of the conventionalized uses are not related to one another: *quedarse quieto* ‘to become calm’ and *quedarse sorprendido* ‘to become surprised’ do not share semantic features. Similarly, *ponerse nervioso* ‘to become nervous’, *ponerse serio* ‘to become serious’, and *ponerse*

pesado ‘to become annoying’ are not closely related semantically. This property suggests that the categories of adjective used with each verb cannot be characterized by discrete features that represent necessary and sufficient conditions for the use of each verb. Rather the usage suggests a representation in the form of local generalizations with conventionalized pairings that are referred to in order to create new pairings. While the data do not lend themselves to description in terms of abstract features that provide strict boundaries to the categories, there are some abstract features that seem to link some uses of the verbs. For instance, a sense of lacking can be found in the clusters centered around *quedarse solo* ‘to end up alone’ and *quedarse quieto, callado, tranquilo* etc. ‘to become still, quiet, tranquil’. All of these indicate the lack of some property such as movement, company, or agitation. We believe that reference to this more abstract notion may unite these categories; however, it is important to note that this feature does not apply to the cluster around *quedarse sorprendido* ‘to be surprised’. Similarly, the data do contain cases where *ponerse* was chosen to express agentivity in the change of state: *ponerse maternal* ‘to act maternal’ for instance, gives the clear impression of a deliberate action. Therefore, some of the features suggested by other authors for characterizing the meanings contributed by the verbs do appear in certain cases. However, the fact that these semantic features do not accompany all uses of a given verb provides important evidence for the nature of the categorization involved.

When we compare *quedarse* and *ponerse* on the one hand with *volverse* and *hacerse* on the other, we see two very different types of productivity: the productive uses of the first two verbs constitute extensions from strong clusters while the productive uses of the second two are more likely to generate dispersed or isolated examples. The first two verbs have higher type and token frequencies, suggesting that the more common type of extension is heavily based on strong exemplar clusters. The second type of extension, where strong models are either not available or not accessed, might appear to constitute a type of ‘default’ construction for the expression of

becoming. However, as observed, even these constructions have some highly entrenched, conventionalized uses that make up the center of exemplar clusters.

Type frequency is a major determinant of productivity in morphology (Bybee 1985, 2001; Baayen and Lieber 1991) and in syntax as well (Goldberg 1995, Bybee and Thompson 1997). In the current case we see an interesting interaction of type frequency with semantic categorization. The two verb constructions with the highest type frequency (which also have the highest token frequency) are those with *quedarse* and *ponerse*; however the types in these two constructions are highly constrained by the meaning of the adjective. One might say that these are the most productive constructions, but that productivity is limited to certain semantic domains. The two constructions with lower type frequency (with adjectives), those with *volverse* and *hacerse*, show productive use but without semantic constraints. A question to ponder is why there are more tokens as well as types with the semantically constrained constructions. Perhaps there are also conventions (especially in literature) about what human changes of state are more likely to be described.

8. The similarity experiment. Although the analyses presented above emerged rather naturally from the corpus data, the fact that only one native speaker was consulted raises the question of whether other speakers would agree with the similarity groupings proposed. In addition, it is useful to determine if the adjectives are judged as similar outside of the change-of-state context. In order to answer these questions, we constructed a questionnaire that asked for similarity judgments between pairs of adjectives and then submitted the results to an analysis using multidimensional scaling. We could not use all 161 adjectives in this experiment because each adjective had to be paired with every other adjective, which would have resulted in more than 10,000 pairs of

adjectives—too many to handle in one experiment. Instead we selected 20 adjectives from the corpus study and elicited similarity judgments for these adjectives.

As mentioned previously, our consultant separated adjectives into a number of groups based on their similarity. We chose twenty adjectives from among the groups that appear in Tables 5c, 6, and 13 as our test cases. That is, we chose fourteen adjectives that occurred with *quedarse* in the corpus: seven adjectives from the cluster around *inmóvil* (including that adjective), and seven from the cluster around *sorprendido* ‘surprised’ (including *sorprendido*) to test these groupings that our consultant had considered highly related semantically. In addition, we chose six adjectives used with *hacerse* which appeared to constitute a much looser semantic organization. All of the twenty adjectives were paired with all of the other adjectives resulting in 190 test items. Since rating 190 pairs of adjectives would be a tiresome and lengthy process for any one person to carry out, we divided the test items into two separate questionnaires containing 95 adjective pairs each. The pairs of adjectives were presented in writing with no context provided; that is, participants were asked to rate each pair in terms of how similar the two words were in meaning without any knowledge that we were interested in what verbs of becoming these adjectives were used with. A five-point scale was used in which five represented ‘highly similar’ and a score of one ‘not similar at all.’

8.1. Participants. The two questionnaires were administered to 77 Spanish-speaking students enrolled at the Universidad de Murcia in Spain. Among the participants, 73 were between the ages of 18 and 29, three were 30 or older, and one did not indicate age; 68 women and 9 men participated. About half of the participants (N = 39) responded to the first questionnaire, and the other half (N = 38) responded to the second.

8.2. Results and discussion. The scores from the questionnaires were inverted for statistical analysis. For example, a score of five, meaning ‘highly similar’ was changed to a score of one meaning ‘closely related.’ In this way, scores reflected the semantic distance between the pairs of adjectives. These data were then submitted to a multidimensional scaling analysis.

Multidimensional scaling is not commonly applied to linguistic analyses so a brief introduction is warranted. Imagine a table containing the distances in numbers of kilometers between all of the capital cities in Europe such as appear in many road atlases. These raw distances do not indicate the relationship of the capitals in terms of the points of the compass; nevertheless, when fed into a multidimensional analysis a two-dimensional map is produced placing the capitals in the relative positions they appear on a map. (See Stalans 1995 for an introduction to multidimensional scaling.)

INSERT FIGURE 8 ABOUT HERE

The mapping of the 20 adjectives in terms of their semantic similarity appears in Figure 8. The proximity in the figure is derived from the scaling and represents how closely related the adjectives were judged to be by the participants. The ovals drawn on the figure represent the semantic grouping we derived earlier from the study of which adjectives occurred with which verb. Remember that the participants’ task was merely to determine semantic similarity; they were not informed of our goal of finding groups of adjectives that take a particular change-of-state verb. Nevertheless, the adjectives taken from Table 13, all of which are normally expressed with *hacerse*, all appear on the right-hand side of the map while adjectives from Table 5c and 6 that generally appear with *quedarse* fall on the left-hand side of the map. What is more, adjectives belonging to the *sorprendido* group (Table 6) cluster in the top left-hand section of the map. Those

that belong to the *inmóvil* group (Table 5c) fall into the bottom left-hand portion of the map. The groupings derived from the experiment, then, strongly support the similarity analysis presented above.

The major way that the experimental results differ from our own analysis is evident in the groupings on the right side of the figure—those used with *hacerse*. From the point of view of a linguist, *bueno* ‘good’, *fuerte* ‘strong’, *rico* ‘rich’ and *famoso* ‘famous’ are not all semantically similar. However, from the point of view of speakers asked to rate how similar the adjectives are, these adjectives may have seemed similar because they all express strong positive values, and some such as *rico* and *famoso* are often paired in discourse. It may be that their use with *hacerse* is in fact related to these contextual pairings, a factor we did not consider in our initial analysis but something that should be considered in future research.

Given these results, we feel confident that most of our groupings would not differ in general form from those resulting from a massive multidimensional scaling study, despite some differences in detail.

Another aspect of our analysis is the central role we have proposed for the higher frequency members of categories. Our hypothesis is that the conventionalized or fixed expressions may serve as an analogical basis for the creation of novel instances of verb + adjective combinations and in this way come to serve as the center for a set of semantically related expressions. In many cases, higher frequency expressions are more semantically general as well, making them good candidates for occupying the center of a category. In the experiment just described, the groupings emerging from the multidimensional analysis show the adjective *sorprendido* as central to a group of adjectives, much as predicted. However, *inmóvil* was also predicted to occupy a central position in its category, which is not, however, where it is found in the figure. As a way of testing more directly the role of token frequency in the formation of these

categories of adjective, we devised a second experiment, which also tests the role of frequency or familiarity in determining acceptability.

19. The acceptability experiment. The second experiment tests the role of token frequency along with semantic similarity in determining category membership. As we have mentioned, an important aspect of an exemplar model is that it represents the token frequency of exemplars directly by strengthening an exemplar with each token of use. We have hypothesized that these stronger exemplars serve as the basis for the production of novel expressions and constitute the central members of the category. Now we take the further step of testing the hypothesis that the usage distributions identified above constitute the basis of the expression of becoming in Spanish. That is, we would like to determine whether or not the experience a language user has with verb + adjective combinations is part of the knowledge of that speaker. In accord with usage-based theory we hypothesize that grammaticality or acceptability judgments are heavily based on familiarity; that is, the speaker's experience with language. Sequences of linguistic units that are of high frequency or resemble sequences of high frequency will be judged more acceptable than those that are of low frequency or do not resemble frequently-used structures. Such correspondences have been found for phonotactic sequences, where subjects give low acceptability ratings to segment combinations that are technically allowed in the language, but of low frequency (Pierrehumbert 1994, Coleman and Pierrehumbert 1997, Vitevitch et al. 1997, Treiman et al. 2000, Bailey and Hahn 2001, Frisch et al. 2001). These studies suggest that grammaticality is based on experience with language and with categorizations of similarity, such as we have proposed in the preceding analysis. For this reason we carried out the following experiment which is designed to determine if the frequency of the exemplars and the kinds of semantic similarity we postulate correlate with acceptability ratings made by native Spanish speakers.

9.1. The questionnaire. The written questionnaire of this experiment consisted of 33 items that the subjects were asked to rate on a Likert-type scale. The subjects were instructed to rate how natural the underlined part of each sentence sounded to them. Underlining was done in order to focus attention on the change-of-state verb and to minimize the influence that other parts of the sentence may have on their judgment. Consider the following test item:

Con el vino y la voz pegajosa se estaba poniendo sentimental.

‘With wine and an over-affectionate voice he was getting sentimental.’

Perfectamente bien

Raro

+-----+-----+-----+-----+

‘Perfectly fine’

‘Odd’

Most of the test sentences were taken directly from the corpora, although some had to be modified slightly in order to fit into the format of the questionnaire. Part of the motivation for the questionnaire was to determine the influence of analogical pull toward the groups with strong clusters. For each of the verbs *quedarse* and *ponerse* we chose three types of test items: (1) items that contained the high-frequency adjectives that make up the core of the exemplar clusters, (2) items containing low-frequency adjectives that are semantically related to one of the core clusters of each verb, (3) low-frequency adjectives that are semantically unrelated to other uses of the verb (see Table 14). The low-frequency test items had only one occurrence in the corpora.

Finding adequate test items proved challenging. For instance, we did not include any test items with the verbs *volverse* and *hacerse* because neither verb had enough high-frequency instances to use for test items to compare with the low-frequency uses. Our exemplar analysis is

also limited to verbs of becoming with animate subjects, but for one test sentence with *quedarse* we had to include a low-frequency adjective that did not appear in our corpus that contained an inanimate subject. The questionnaire items were randomized and presented to the participants in paper and pencil format. Their task was to circle a '+' on the five-point scale that ranged from 'perfectly fine' (1) to 'odd' (5). No time limit was imposed.

INSERT TABLE 14 ABOUT HERE

9.2. Participants. A total of 48 native Spanish speakers participated in the experiment, all of whom were students enrolled at the Universidad de Murcia in Spain.

9.3. Results and discussion. The responses on the questionnaire are shown in Table 15 and schematized in Figures 9 and 10. A Chi square analysis of the 'perfectly fine' responses (number 1 on the graph) indicates that the differences in their distribution are highly significant (*quedarse* $X^2(1) = 51.4, p < .0001$; *ponerse* $X^2(1) = 32.3, p < .0001$). Most participants judged the high-frequency items on the 'perfectly fine' side of the scale. The same was true of the low-frequency items that are semantically similar to the high-frequency clusters. We compared the high- and low-frequency 'perfectly fine' responses for both verbs together, and found a significant difference based on frequency ($X^2(1) = 6.22, p < .013$). The strongest result is that the low-frequency items that are not related to high-frequency clusters were judged to be much less acceptable than the low-frequency related items (*quedarse* $X^2(1) = 32.9, p < 0.0001$; *ponerse* $X^2(1) = 16.0, p < 0.0001$). The responses showed that both semantic similarity and token frequency were significant in determining the acceptability of the pairings of verb + adjective.

INSERT TABLE 15 ABOUT HERE

INSERT FIGURE 9 ABOUT HERE

INSERT FIGURE 10 ABOUT HERE

The results show that frequency, together with similarity to exemplars experienced by language users, influences acceptability judgments. The fact that high-frequency exemplars of constructions are given high acceptability ratings indicates that frequency of occurrence is not only reflected in the mind of the participants, but also serves as an important factor in determining intuitions of acceptability. It is important to remember that almost all the sentences presented in the questionnaire had actually occurred in the corpus; indeed, they were all possible sentences of Spanish. Nonetheless, the subjects identified significant differences in their acceptability. The high-frequency expressions and the low-frequency expressions that were semantically related to the high-frequency ones were judged more acceptable than the isolated (low-frequency unrelated) combinations. This result indirectly confirms the centrality of the higher frequency expressions in forming similarity clusters, as both the high-frequency and the lower-frequency related expressions were judged to be more acceptable than the lower-frequency isolated expressions.

We can further conclude that acceptability is a function of experience: those exemplars that have been experienced frequently and exemplars that are similar to them in meaning and structure are judged to be more acceptable. This does not mean that a new sentence that has never been experienced before will be judged as unacceptable; rather its acceptability will be assessed on the basis of its similarity to sentences already experienced. Since an adult speaker has experienced a vast number of sentences on which to base these judgments, even a semantically anomalous sentence such as *Colorless green ideas sleep furiously* can be recognized as using the words of English as well as the grammatical structures familiar from other sentences.

Most theories of prototype categorization provide no role for token frequency in determining the central member of the category. However, studies by Nosofsky (1988) have shown that changing the frequency of exemplars in experience can change the center and boundaries of a category. Besides agreeing with Nosofsky's findings, our study concurs with a large body of psycholinguistic evidence demonstrating frequency effects in language processing (e.g. Hooper 1976, Scarborough et al. 1977, Balota and Chumbley 1984, Allen et al. 1992).

10. Conclusion. This study was intended as a contribution to the evolving understanding of constructions and how instances of specific construction relate to one another. We focused primarily on the categorization that occurs in the open slots in constructions, arguing for an exemplar model for these categories. For our analysis of the choice of 'become' verbs with different adjectives in Spanish, we chose to study the actually occurring examples in a large corpus. The analysis was exhaustive in the sense that we took into account every example that had an animate subject and an adjective (or prepositional phrase) with the four verbs. We also attended to the frequency of use of the particular exemplars. This study, in contrast to those that use a limited number of invented sentences, provided more dimensions upon which to found an analysis. The resulting categorization, which is based on similarity between individual adjectives, is more finely textured than an analysis that seeks broad abstract features. A majority of the tokens fit nicely into radial categories with high-frequency exemplars at their center and in this sense strongly resemble natural categorization as evidenced in other domains of language, such as in the meanings of particular words and changes they undergo (Lakoff 1987, Geeraerts 1997) and in categorization of non-linguistic stimuli, such as colors (Nosofsky 1988). The open slots in constructions are subject to the same categorization processes as other cultural and natural objects.

The relative frequencies and the semantic similarities found in the corpus material point directly to the analysis proposed.

The two experiments confirmed the role of frequency and semantic similarity in determining the form of the subjects' cognitive representations or grammar. The semantic similarity experiment showed that along with shared semantic features such as we used in our analysis, subjects used metonymic relations to group adjectives. The results of the second experiment showed that judgments of acceptability are sensitive both to semantic similarity and frequency of use, suggesting that grammatical knowledge is built up by registering and categorizing experience with language. The study demonstrates that grammar can be viewed as a cognitive map of linguistic experiences, with acceptability judgments that are based on familiarity, and the productive or creative use of language based on reference to categorized instances of prior use.

References

- Allen, Phillip, Michelle McNeal, and Donna Kvak. 1992. Perhaps the lexicon is coded as a function of word frequency. *Journal of Memory and Language* 31.826-844.
- Baayen, Harald, and Rochelle Lieber. 1991. Productivity and English derivation: a corpus-based study. *Linguistics* 29.801-843.
- Bailey, Todd, and Ulrich Hahn. 2001. Determinants of wordlikeness: phonotactics or lexical neighborhoods? *Journal of Memory and Language* 44.568-591.
- Balota, David A., and James I. Chumbley. 1984. Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected access stage. *Journal of Experimental Psychology: Human Perception and Performance* 10.340-357.
- Bod, Rens. 1998. *Beyond grammar*. Stanford, CA: Center for the Study of Language and Information.
- Boylard, Joyce Tang. 1997. *Morphosyntactic change in progress: a psycholinguistic approach*. Berkeley: University of California Doctoral Dissertation.
- Bybee, Joan. 1985. *Morphology: a study of the relation between meaning and form*. Philadelphia: John Benjamins.
- Bybee, Joan. 1998. The emergent lexicon. *Chicago Linguistic Society 34: The Panels*, ed. by M. Catherine Gruber, Derrick Higgins, Kenneth Olson, and Tamra Wysocki, 421-435. University of Chicago: Chicago Linguistic Society.
- Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.
- Bybee, Joan. 2003. Mechanisms of change in grammaticization: the role of frequency. *The Handbook of Historical Linguistics*, ed. by Brian D. Joseph and Richard D. Janda, 602-623. Oxford: Blackwell Publishers.

- Bybee, Joan. 2006. From usage to grammar: the mind's response to repetition. *Language* 82.
- Bybee, Joan, and Paul Hopper. (eds.) 2001. *Frequency and the emergence of linguistic structure*. Amsterdam, Philadelphia: John Benjamins.
- Bybee, Joan, and Carol L. Moder. 1983. Morphological classes as natural categories. *Language* 59.251-270.
- Bybee, Joan, and William Pagliuca. 1987. The evolution of future meaning. *Papers from the VIIIth international conference on historical linguistics*, ed. by Anna Giacalone Ramat, Onofrio Carruba, and Giuliano Bernini, 109-122. Amsterdam: John Benjamins.
- Bybee, Joan, and Sandra Thompson. 1997. Three frequency effects in syntax. *Berkeley Linguistic Society* 23.65-85.
- Chandler, Steve. 2002. Skousen's analogical approach as an exemplar-based model of categorization. *Analogical modeling*, ed. by Royal Skousen, Deryle Lonsdale, and Dilworth B. Parkinson, 51-105. Amsterdam: John Benjamins.
- Coleman, John, and Janet Pierrehumbert. 1997. Stochastic phonological grammars and acceptability. *Computational phonology: proceedings of the third meeting of the Association for Computational Linguistics special interest group in computational phonology*, 49-56. Somerset, NJ: Association for Computational Linguistics.
- Coste, Jean, and Augustin Redondo. 1965. *Syntaxe de l'espagnol moderne*. Paris: Société d'Édition d'Enseignement Supérieur.
- Crespo, Luis A. 1949. To become. *Hispania* 32.210-12.
- Croft, William. 2001. *Radical construction grammar*. Oxford: Oxford University Press.
- Culicover, Peter W. 1999. *Syntactic nuts: hard cases, syntactic theory and language acquisition*. Oxford: Oxford University Press.

- Eberenz, Rolf. 1985. Aproximación estructural a los verbos de cambio en Iberorromance. *Linguistique comparée et typologie des langues romanes*, ed. by Jean-Claude Bouvier, 460-475. Aix en Provence: Université de Provence.
- Eddington, David. 1999. On 'becoming' in Spanish: a corpus analysis of verbs of expressing change of state. *Southwest Journal of Linguistics* 18.23-46.
- Eddington, David. 2002. Disambiguating Spanish change of state verbs. *Hispania* 85.921-929.
- Erman, Britt, and Beatrice Warren. 2000. The idiom principle and the open choice principle. *Text* 20.29-62.
- Fente, R. 1970. Sobre los verbos de cambio o devenir. *Filología Moderna* 38.157-72.
- Fillmore, Charles J., Paul Kay, and Mary C. O'Connor. 1988. Regularity and idiomatity in grammatical constructions. *Language* 64.501-538.
- Frisch, Stefan, Nathan R. Large, Bushra Zawaydeh, and David B. Pisoni. 2001. Emergent phonotactic generalizations in English and Arabic. *Frequency and the emergence of linguistic structure*, ed. by Joan Bybee and Paul Hopper, 159-180. Amsterdam: John Benjamins.
- Geeraerts, Dirk. 1997. *Diachronic prototype semantics: a contribution to historical lexicology*. Oxford: Oxford University Press.
- Goldberg, Adele. 1995. *Constructions: a construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goldberg, Adele E. 2003. Constructions: a new theoretical approach to language. *Trends in Cognitive Science* 7.219-224.
- Hay, Jennifer. 2001. Lexical frequency in morphology: is everything relative? *Linguistics* 39.1041-1070.

- Hooper, Joan B. 1976. Word frequency in lexical diffusion and the source of morphophonological change. *Current progress in historical linguistics*, ed. by William Christie, 96-105. Amsterdam: North Holland.
- Hopper, Paul. 1991. On some principles of grammaticization. *Approaches to grammaticalization*, vol. I, ed. by Elizabeth C. Traugott and Bernd Heine, 17-35. Amsterdam: John Benjamins.
- Israel, Michael. 1996. The way constructions grow. *Conceptual structure, discourse and language*, ed. by Adele E. Goldberg, 217-230. Stanford: CSLI Publications.
- Jespersen, Otto. 1942. *A modern English grammar on historical principles, VI: Morphology*. Copenhagen: Munksgaard.
- Johnson, Keith. 1997. Speech perception without speaker normalization. *Talker variability in speech processing*, ed. by Keith Johnson and John W. Mullennix, 145-65. San Diego: Academic Press.
- Lakoff, George. 1987. *Women, fire and dangerous things*. Chicago: University of Chicago Press.
- Langacker, Ronald. 1987. *Foundations of cognitive grammar, volume 1: theoretical prerequisites*. Stanford: Stanford University Press.
- Marcos Marín, Francisco. 1992. Corpus oral de referencia del español contemporáneo. Textual corpus, Universidad Autónoma de Madrid.
- Medin, Douglas L., and Marguerite M. Schaffer. 1978. Context theory of classification learning. *Psychological Review* 85.207-238.
- Murphy, Gregory L. 2002. *The big book of concepts*. Cambridge, MA: MIT Press.
- Nosofsky, Robert M. 1988. Similarity, frequency, and category representation. *Journal of Experimental Psychology: Learning, Memory and Cognition* 14.54-65.

- Pawley, Andrew, and Frances Hodgetts Syder. 1983. Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *Language and communication*, ed. by Jack C. Richards and Richard W. Schmidt, 191-226. Longman: London.
- Pierrehumbert, Janet. 1994. Syllable structure and word structure: a study of triconsonantal clusters in English. *Phonological structure and phonetic form. Papers in laboratory phonology III*, ed. by Patricia Keating, 168-190. Cambridge: Cambridge University Press.
- Pierrehumbert, Janet. 2001. Exemplar dynamics: word frequency, lenition and contrast. *Frequency and the emergence of linguistic structure*, ed. by Joan Bybee and Paul Hopper, 137-158. Amsterdam: John Benjamins.
- Pierrehumbert, Janet. 2002. Word-specific phonetics. In *Laboratory Phonology 7*, ed. by Carlos Gussenhoven and Natasha Warner, 101-139. Berlin: Mouton de Gruyter.
- Rosch, Eleanor, and Carolyn Mervis. 1975. Family resemblances: studies in the internal structure of categories. *Cognitive Psychology* 8.382-439.
- Scarborough, Don. L., Charles Cortese, and Hollis S. Scarborough. 1977. Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance* 3.1-17.
- Sinclair, John. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Skousen, Royal. 1989. *Analogical modeling of language*. Dordrecht: Kluwer Academic.
- Stalans, Loretta J. 1995. Multidimensional scaling. *Reading and understanding multivariate statistics*, ed. by Laurence G. Grimm and Paul R. Yarnold, 137-168. Washington D.C.: American Psychological Association.
- Sweet, Henry. 1882. [1965]. *Anglo-Saxon primer*, 9th edition. Oxford: Clarendon Press.
- Taylor, John R. 1995. *Linguistic categorization*. Oxford: Oxford University Press.

- Treiman, Rebecca, Brett Kessler, Stephanie Knewasser, and Ruth Tincoff. 2000. Adults' sensitivity to phonotactic probabilities. *Papers in laboratory phonology V: Acquisition and the lexicon*, ed. by Michael Broe and Janet Pierrehumbert, 269-282. Cambridge: Cambridge University Press.
- Vitevitch, Michael S., Paul A. Luce, Jan Charles-Luce, and David Kemmerer. 1997. Phonotactics and syllable stress: implications for the processing of spoken nonsense words. *Language and Speech* 40.47-62.
- Wittgenstein, Ludwig. 1953. *Philosophical investigations*. (Anscombe, G.E.M., trans.). Oxford: Basil Blackwell.
- Wray, Alison. 2002. *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Notes

¹ Other ways of expressing ‘entering into a state’ are available in Spanish. A somewhat productive derivational process converts adjectives into inchoative verbs: *rico* ‘rich’ is the basis of *enriquecerse* ‘to get rich’; *viejo* ‘old’ gives *envejecerse* ‘to get old’; *flaco* ‘thin’ gives *enflaquecerse* ‘to grow thin.’ This process and its functions will not be discussed in this paper.

² Pierrehumbert (2001) proposes that a token is mapped onto an existing phonetic exemplar if there are no ‘just noticeable differences’ between the token and the existing exemplar.

³ One effect of extremely high token frequency is to render an exemplar of a construction autonomous from other exemplars or other related items (Bybee 1985, Bailey and Hahn 2001, Hay 2001). It appears that the degree of token frequency found in the data to be analyzed here is not sufficient to cause autonomy, as we find evidence of relatedness in all the cases of high-frequency exemplars.

⁴ If by this definition Goldberg means that single morphemes can be constructions, we disagree. A morpheme will occur in a construction—indeed, most morphemes are defined by the constructions they appear in—but a morpheme alone is not a construction.

⁵ The ‘total types’ will be less than the total of types in both corpora because some of the same types occurred in both the written and spoken corpora.

⁶ Fixed expression (294 cases found on Google).

⁷ *Ponga* is the Present Subjunctive form of *poner*.

⁸ *Puesto* is the Past Participle of *poner*.

⁹ *Puso* is the 3s Preterite form of *poner*.

¹⁰ We also extracted and analyzed the uses of these verbs with inanimate subjects. Because we found so little overlap between the meanings of the adjectives used with animate and inanimate subjects we decided to analyze only the uses with the animate subjects. Thus while there may be

more coherent semantic clustering in some cases with inanimate subjects, it does not change the analysis for the animate subjects.

	spoken		written		total
	tokens	types	tokens	types	types ⁵
<i>quedarse</i>	68	40	181	54	69
<i>ponerse</i>	36	23	85	45	62
<i>hacerse</i>	8	8	16	11	16
<i>volverse</i>	7	2	22	13	14
total	119		304		161

Table 1

Number of occurrences and number of types with a human subject and an adjective in a spoken corpus of 1.1 million words and a written corpus of about one million words (Marcos Marín 1992, Eddington 1999).

adjective	quedarse	ponerse	volverse	hacerse
<i>loco</i> 'crazy'	1		16	
<i>triste</i> 'sad'	5	2		
<i>contento</i> 'happy'	1	4		
<i>serio</i> 'serious'	1	3		
<i>ciego</i> 'blind'	1	3		
<i>tieso</i> 'rigid, stiff'	1	1		
<i>viejo</i> 'old'	1	1	1	1
<i>pesado</i> 'annoying'		6	1	
<i>negro</i> 'black'		1	1	
<i>cursi</i> 'tacky'		1		1
<i>duro</i> 'hard'	2			1
<i>lívido</i> 'pale'	1	1		

Table 2

Adjectives used with more than one verb.

adjective	spoken	written
<i>solo</i> 'alone'	7	21
<i>soltera</i> 'single, unmarried'	1	2
<i>aislado</i> 'isolated'	2	0
<i>a solas</i> 'alone'	1	0
<i>sin novia</i> 'without a girlfriend'	1	0
opposite:		
<i>emparejado</i> 'paired with'	1	0

Table 3

Number of adjectives related to *solo* used with *quedarse* in the spoken and written corpus.

adjective	spoken	written
<i>quieto</i> 'calm, still'	0	29
<i>dormido</i> 'asleep'	1	27
<i>inmóvil</i> 'motionless'	0	17
<i>tranquilo</i> 'calm, peaceful'	6	10
<i>callado</i> 'quiet, not speaking'	0	8
<i>a gusto</i> 'pleased'	3	1

Table 4

The central members of the *quieto* cluster of expressions used with *quedarse*.

adjective	spoken	written
<i>quieto</i> 'calm, still'	0	29
<i>pensativo</i> 'thoughtful'	0	2
<i>serio</i> 'serious'	0	1

Table 5a

Adjectives used with *quedarse* and grouped with *quieto* that are not grouped with any other central member.

adjective	spoken	written
<i>dormido</i> 'asleep'	1	27
<i>muerto</i> 'dead'	0	1
<i>aletargado</i> 'drowsy'	0	1
opposites:		
<i>despierto</i> 'awake'	0	1
<i>atento</i> 'attentive'	0	1

Table 5b

Adjectives used with *quedarse* that are grouped with *dormido*.

adjective	spoken	written
<i>inmóvil</i> 'motionless'	0	17
<i>parado</i> 'stopped, standing'	2	0
<i>tieso</i> 'stiff'	0	3
<i>duro</i> 'hard'	0	2
<i>petrificado</i> 'turned to stone'	0	1
<i>de piedra</i> 'made of stone'	1	0
<i>paralizado</i> 'paralyzed'	0	1
<i>seca</i> 'dry'	0	1
<i>clavado al suelo</i> 'nailed to the ground'	0	1
<i>inoperante</i> 'inoperative'	1	0
<i>encerrado</i> 'closed in'	0	1

Table 5c

Adjectives used with *quedarse* that are grouped with *inmóvil* indicating physical motionlessness.

adjectives	spoken	written
<i>inmóvil</i> 'motionless'	0	17
<i>trabado</i> 'stuck, jammed'	0	2
<i>atrapado</i> 'trapped'	0	1
<i>atollado</i> 'in a jam'	0	1
<i>pegado</i> 'stuck'	0	1
<i>enmarado en lo inmóvil</i> 'becalmed'	0	1

Table 5d

Adjectives used with *quedarse* that are grouped with *inmóvil* which extend motionlessness metaphorically.

adjectives	spoken	written
<i>tranquilo</i> 'calm, peaceful'	6	10
<i>conforme</i> 'satisfied'	1	4
<i>satisfecho</i> 'satisfied'	2	1
<i>relajadito</i> 'relaxed'	1	0
<i>mansito</i> 'meek, gentle'	0	1
opposite:		
<i>insatisfecho</i> 'dissatisfied'	0	1

Table 5e

Adjectives used with *quedarse* that are grouped with *tranquilo*.

adjective	spoken	written
<i>callado</i> 'quiet, not speaking'	0	8
<i>silencioso</i> 'silent'	0	1
<i>mudo</i> 'not speaking'	0	4

Table 5f

Adjectives used with *quedarse* that are grouped with *callado*.

adjective	spoken	written
<i>a gusto</i> 'pleased'	3	1
<i>contento</i> 'happy'	1	0
more distant:		
<i>orgullosísimo</i> 'very proud'	1	0

Table 5g

Adjectives used with *quedarse* that are grouped with *a gusto*.

adjective	spoken	written
<i>sorprendido</i> 'surprised'	4	3
<i>deshumbrado</i> 'dazzled'	1	0
<i>fascinado</i> 'fascinated'	0	1
<i>asombrado</i> 'amazed'	0	1
<i>asustado</i> 'frightened'	1	0
<i>seco</i> 'dry, frightened'	1	0
<i>acojonado</i> 'impressed'	1	0
<i>trastornado</i> 'disturbed'	0	1
<i>alucinado</i> 'amazed'	3	0
<i>loco</i> 'crazy'	1	0
<i>frío</i> 'surprised'	1	1
<i>perplejo</i> 'perplexed'	0	1
<i>pasmado</i> 'stunned'	0	1
<i>estupefacto</i> 'stupefied'	0	2
<i>atónito</i> 'astounded'	0	1
<i>preocupado</i> 'worried'	0	1
<i>frustrado</i> 'frustrated'	1	0
<i>colgado</i> 'disappointed'	1	0

Table 6

Adjectives used with *quedarse* that are grouped with *sorprendido*.

adjective	spoken	written
<i>triste</i> 'sad'	1	4
<i>nostálgico</i> 'nostalgic'	1	0
<i>hecho polvo</i> 'depressed, fatigued'	1	0
<i>como un estropajo</i> 'like a scouring pad'	1	0

Table 7

Expressions with *quedarse* that are grouped with *triste*.

adjective	spoken	written
<i>embarazada</i> 'pregnant'	4	0
<i>preñada</i> 'pregnant'	3	1
related:		
<i>desnutrido</i> 'malnourished'	1	0
<i>en bolas</i> 'naked'	1	0
<i>ciego</i> 'blind'	0	4
<i>asfixiado</i> 'suffocated'	1	0
<i>calvo</i> 'bald'	2	0
<i>encogido</i> 'cringing'	0	1
<i>mejor</i> 'better'	1	0
<i>viejo</i> 'old'	1	0
<i>pelado</i> 'shaved'	0	1
<i>toruno</i> 'castrated'	0	1
<i>delgado</i> 'thin'	0	1
<i>estéril</i> 'sterile'	0	1

Table 8

Adjectives used with *quedarse* indicating physical states.

adjective	spoken	written
<i>nervioso</i> 'nervous'	7	10
<i>lívido</i> 'pale, purple'	1	0
<i>pálido</i> 'pale, pallid'	0	6
<i>blanco</i> 'pale, pallid'	0	1
<i>rígido</i> 'rigid'	0	6
<i>tieso</i> 'stiff'	0	1

Table 9a

Adjectives grouped with *ponerse nervioso*.

adjective	spoken	written
<i>furioso</i> 'furious'	2	3
<i>histérico</i> 'hysterical'	1	3
<i>agresivo</i> 'aggressive'	0	1
<i>bravo</i> 'angry, aggressive'	0	2
<i>enojadísimo</i> 'very angry'	0	1

Table 9b

Adjectives grouped with *ponerse furioso* and *ponerse histérico*.

adjective	spoken	written
<i>rojo</i> 'red'	0	2
<i>colorado</i> 'red'	0	3
<i>encarnado</i> 'red'	0	2
<i>como tomate</i> 'like a tomato'	0	1

Table 9c

Adjectives indicating turning red in the face; related to *ponerse nervioso*.

adjective	spoken	written
<i>serio</i> 'serious'	1	2
<i>triste</i> 'sad'	0	2
<i>grave</i> 'grave'	0	1
<i>solemne</i> 'solemn'	1	0
<i>trágico</i> 'tragic'	1	0
<i>gótico</i> 'Gothic'	1	0

Table 10

Adjectives grouped with *ponerse serio*.

adjective	spoken	written
<i>pesado</i> 'annoying'	3	3
<i>cargoso</i> 'annoying'	0	1
<i>fastidioso</i> 'annoying'	0	1
<i>molesto</i> 'annoying'	0	1
<i>inaguantable</i> 'intolerable'	0	1
<i>borde</i> 'nasty'	1	0
<i>tozudo</i> 'stubborn'	1	0
<i>insolente</i> 'rude'	0	1
<i>tonto</i> 'silly'	1	0
<i>bobo</i> 'silly'	1	0
<i>cursi</i> 'tacky, affected'	1	0
<i>alzado</i> 'big-headed'	0	1
<i>ácido</i> 'sour'	1	0

Table 11

Adjectives grouped with *ponerse pesado*.

adjective	spoken	written
<i>loco</i> 'crazy'	6	10
<i>idiota</i> 'idiotic'	0	1
<i>llena de furia</i> 'full of fury'	0	1
<i>mística</i> 'mystical'	0	1
<i>pesado</i> 'annoying'	0	1
<i>raquitico</i> 'rickety, weak'	0	1
<i>fino</i> 'fine'	0	1
<i>exquisito</i> 'exquisite'	1	0
<i>esquivo</i> 'shy'	0	1
<i>ensimismado</i> 'introverted'	0	1
<i>sumiso</i> 'submissive'	0	1
<i>susceptible</i> 'susceptible'	0	1
<i>mieles</i> 'sweet (lit. honeys)'	0	1
<i>negro</i> 'black'	0	1
<i>viejo</i> 'old'	0	1

Table 12

Adjectives used with *volverse*.

adjective	spoken	written
<i>aburrido</i> 'boring'	1	0
<i>cursi</i> 'tacky'	1	0
<i>consciente</i> 'aware of'	1	2
<i>realista</i> 'realistic'	0	1
<i>responsable</i> 'responsible'	0	1
<i>mayor</i> 'grown up'	1	1
<i>viejo</i> 'old'	0	2
<i>duro</i> 'hard'	0	1
<i>fuerte</i> 'strong'	1	1
<i>invulnerable</i> 'invulnerable'	0	1
<i>no inferior</i> 'not inferior'	1	0
<i>digno</i> 'dignified'	0	1
<i>bueno</i> 'good'	1	0
<i>famoso</i> 'famous'	1	1
<i>rico</i> 'rich'	0	2
<i>visible</i> 'visible'	0	2
<i>presente</i> 'present'	0	3

Table 13

Adjectives used with *hacerse*.

	High-frequency	Low-frequency Related	Low-frequency Unrelated
<i>Quedarse</i>	<i>quietos</i> 'still' <i>dormida</i> 'asleep' <i>sola</i> 'alone' <i>sorprendida</i> 'surprised'	<i>paralizada</i> 'paralyzed' <i>pasmados</i> 'shocked' <i>perplejo</i> 'perplexed' <i>clavado</i> 'nailed'	<i>orgullosísimo</i> 'proud' <i>desnutrido</i> 'malnourished' <i>convencida</i> 'convinced' <i>redondo</i> 'round'
<i>Ponerse</i>	<i>nervioso</i> 'nervious' <i>furioso</i> 'furious' <i>enferma</i> 'sick' <i>pesado</i> 'annoying'	<i>inaguantable</i> 'intolerable' <i>negro</i> 'black' <i>agresivo</i> 'aggressive' <i>revoltoso</i> 'rebellious'	<i>viejo</i> 'old' <i>maternal</i> 'maternal' <i>sentimental</i> 'sentimental' <i>putona</i> 'slutty'

Table 14

Verbs and adjectives used in the second experiment.

	'perfectly fine'			'odd'		no response
	1	2	3	4	5	
<i>quedarse</i>						
high-frequency	111	41	16	15	9	0
low-frequency related	87	47	41	6	8	3
low-frequency unrelated	26	38	39	42	46	1
<i>ponerse</i>						
high-frequency	115	43	15	11	8	0
low-frequency related	89	38	32	22	10	1
low-frequency unrelated	43	48	35	27	38	1

Table 15

Responses to the questionnaire.

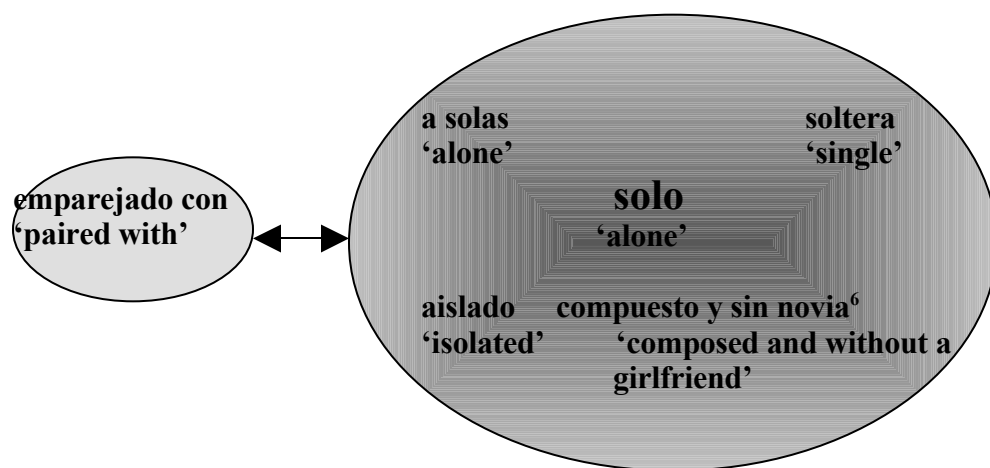
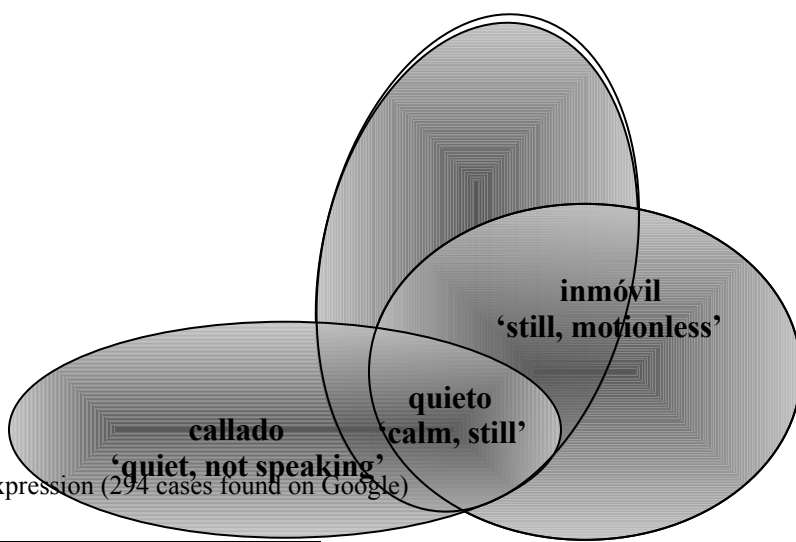


Figure 1

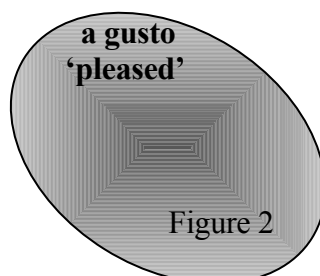
Clusters centering on *quedarse solo*.



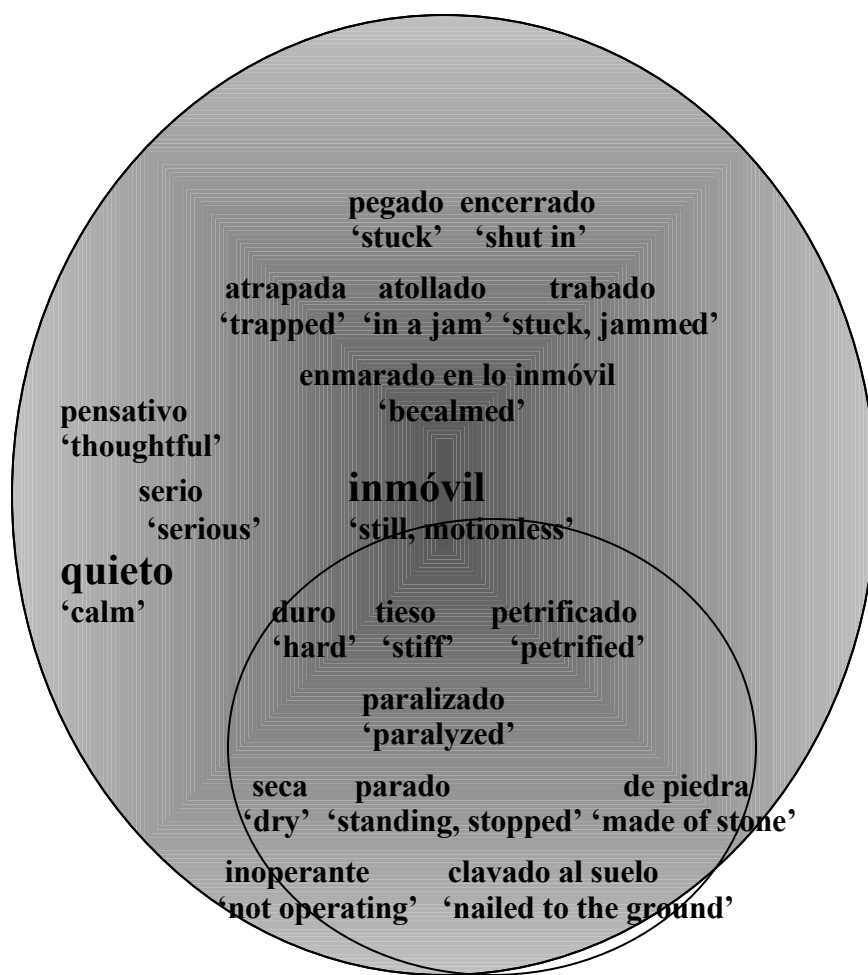
⁶idiomatic expression (294 cases found on Google)

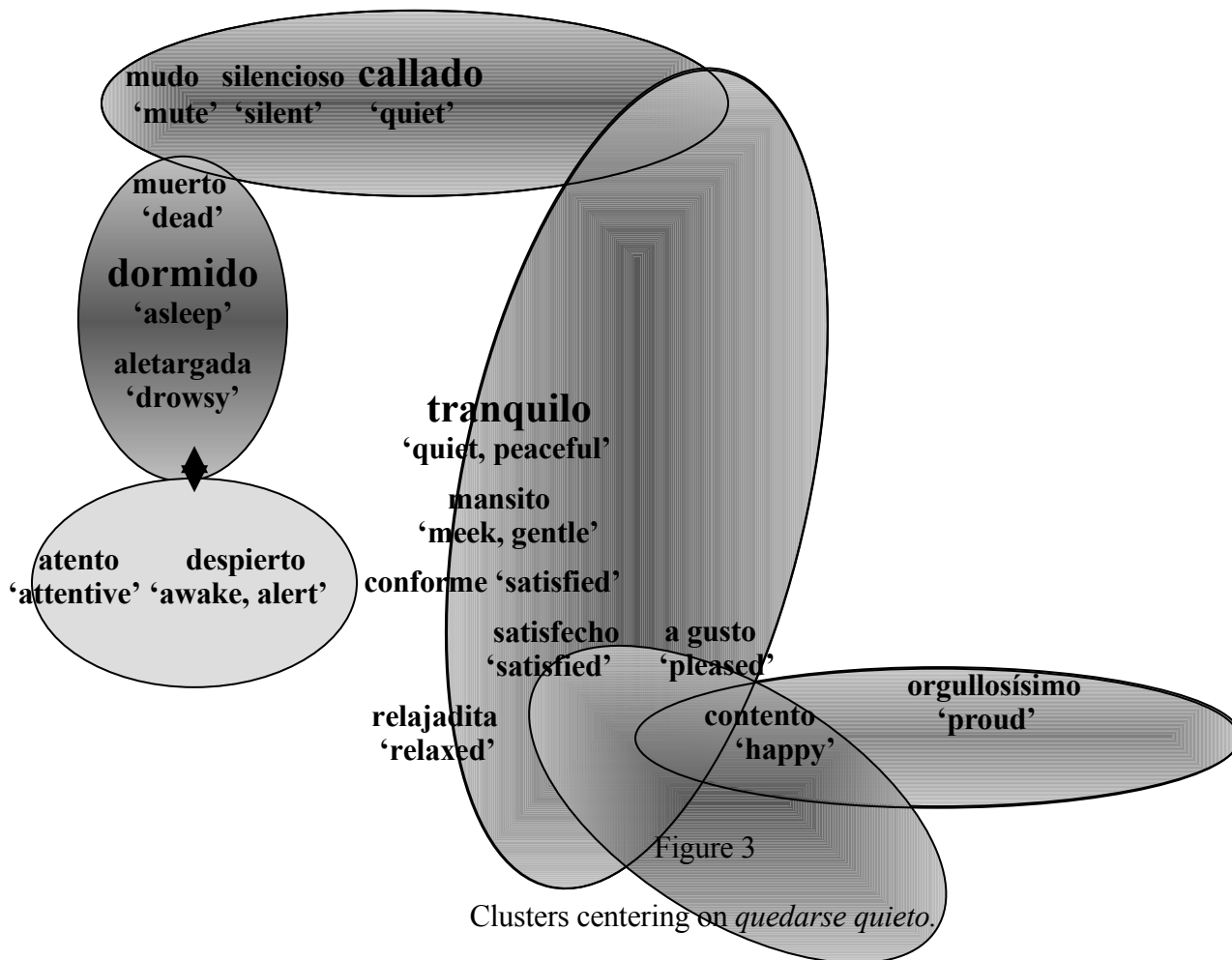


tranquilo
'quiet, peaceful'



Central members of clusters centering on *quedarse quieto*.





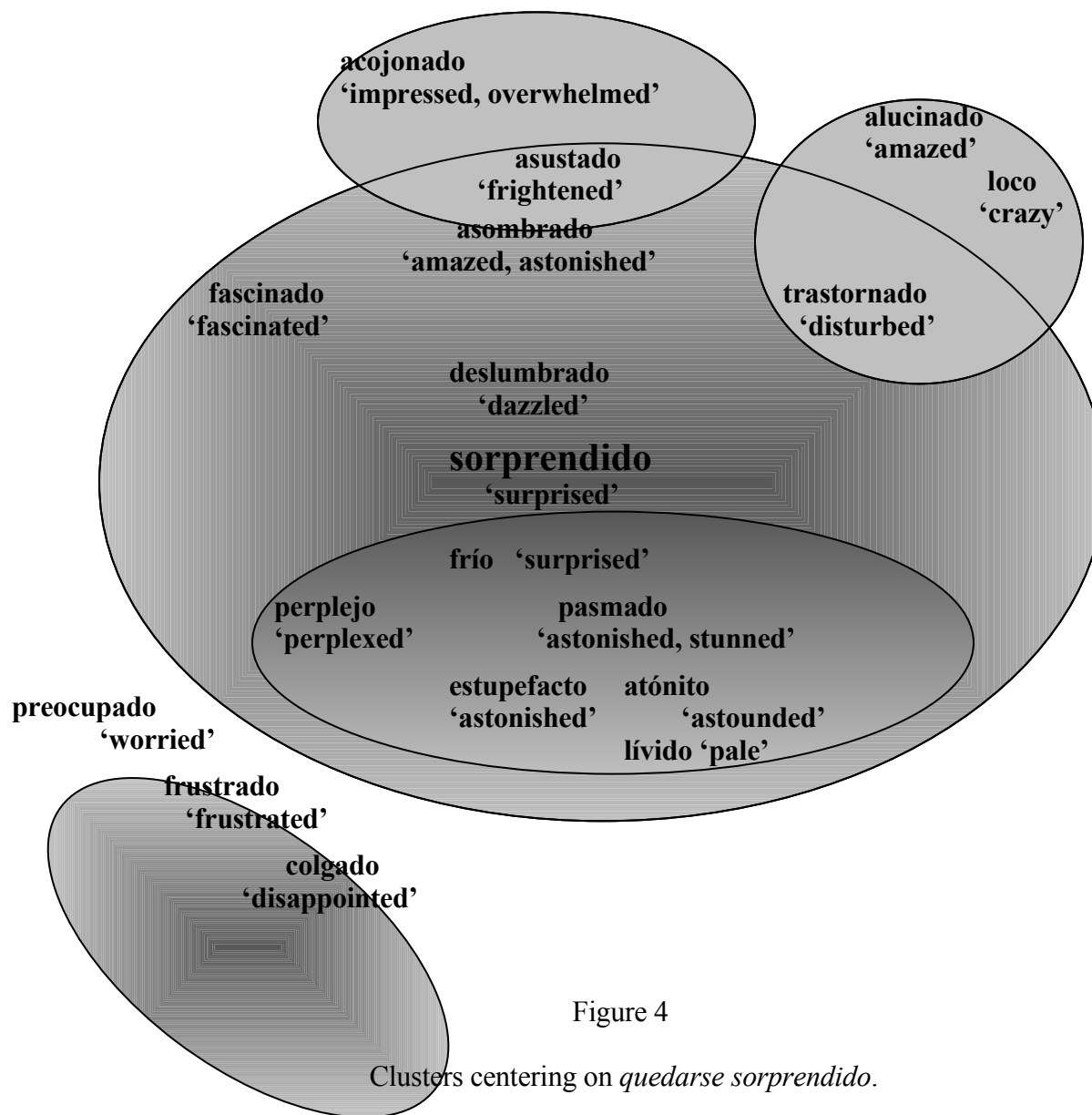


Figure 4

Clusters centering on *quedarse sorprendido*.



Figure 5

Cluster centering on *quedarse triste*.

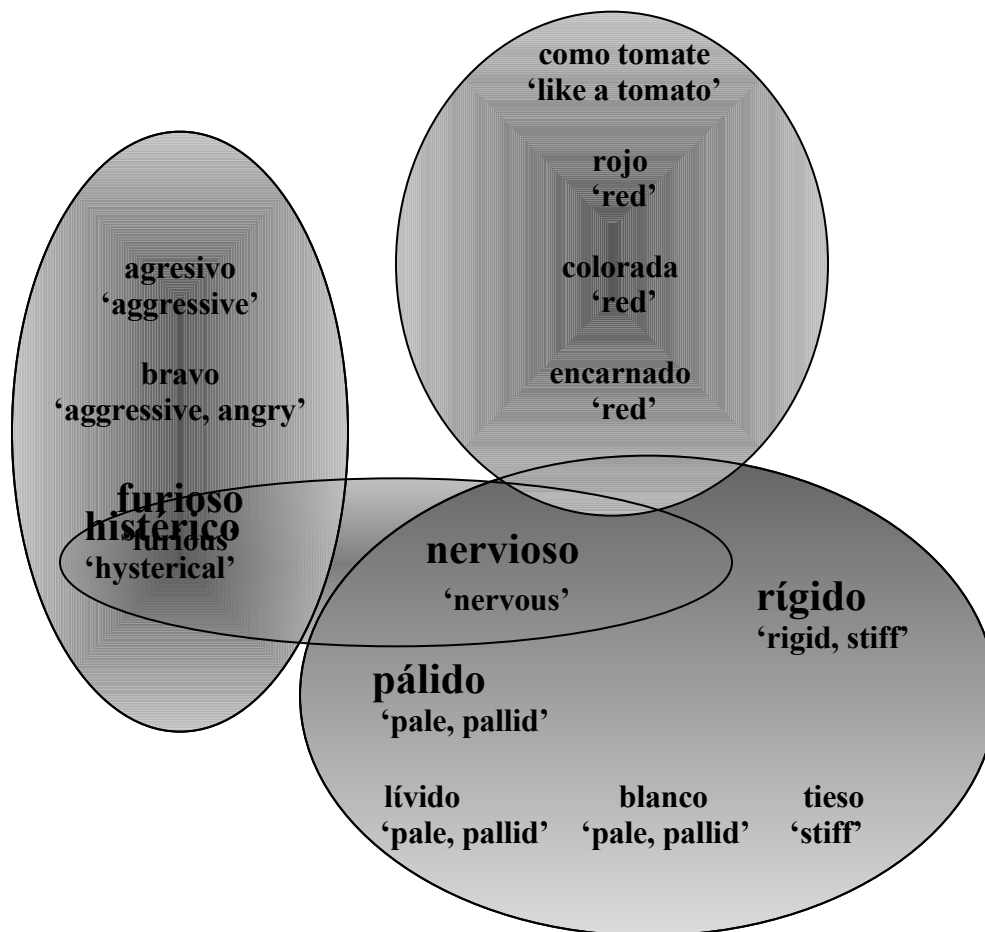


Figure 6

Clusters centering on *ponerse nervioso*.

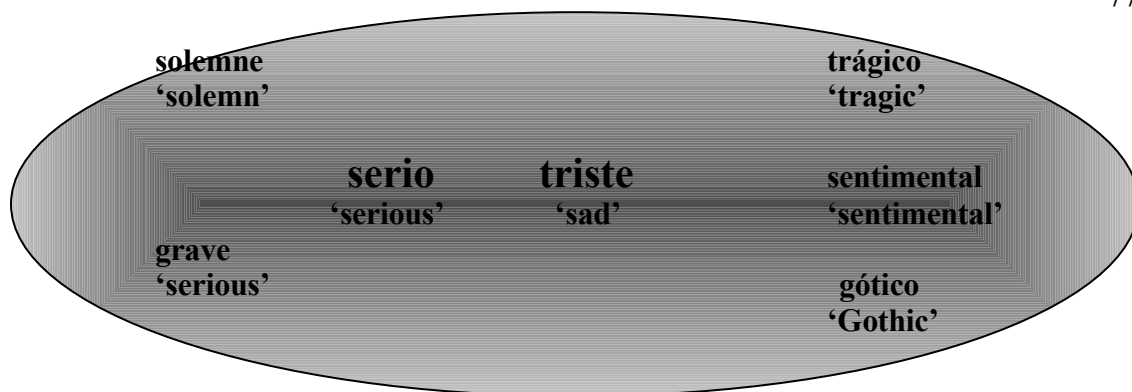


Figure 7

Cluster centering on *ponerse serio/triste*.

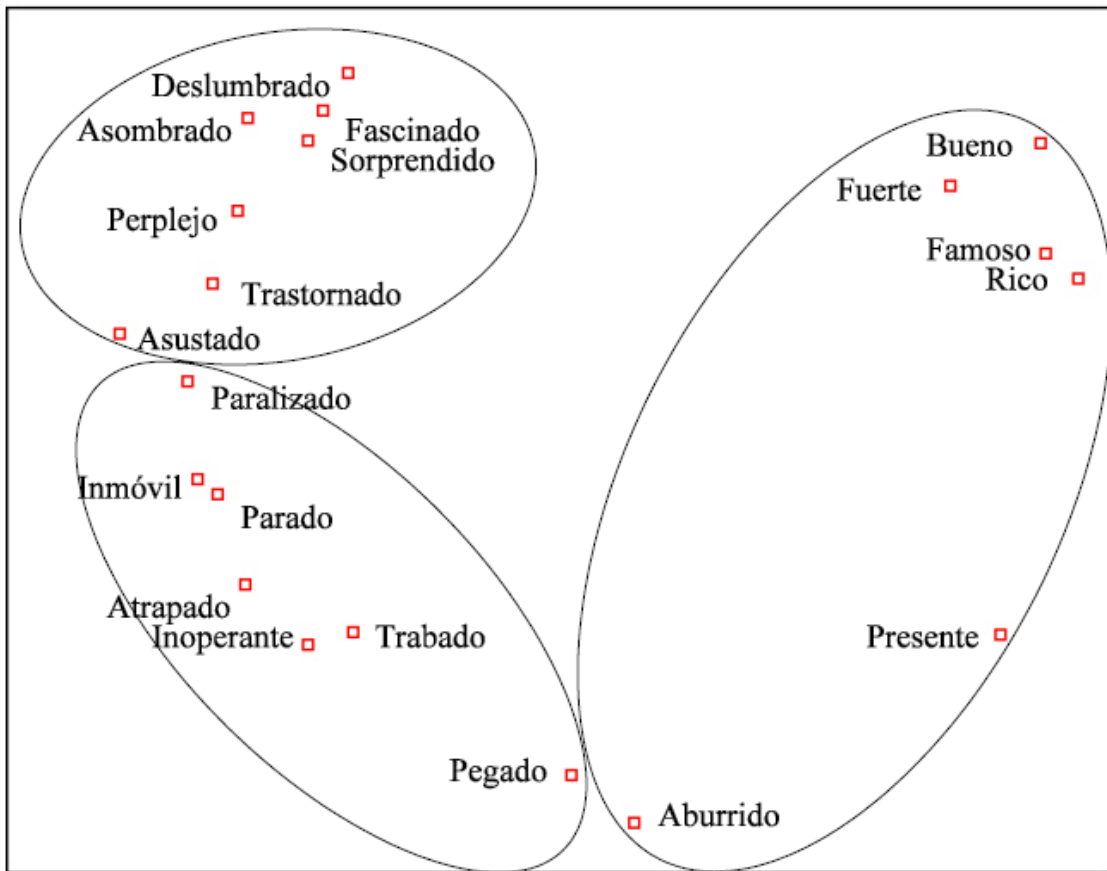


Figure 8

Results of multidimensional scaling of ratings from the similarity experiment.

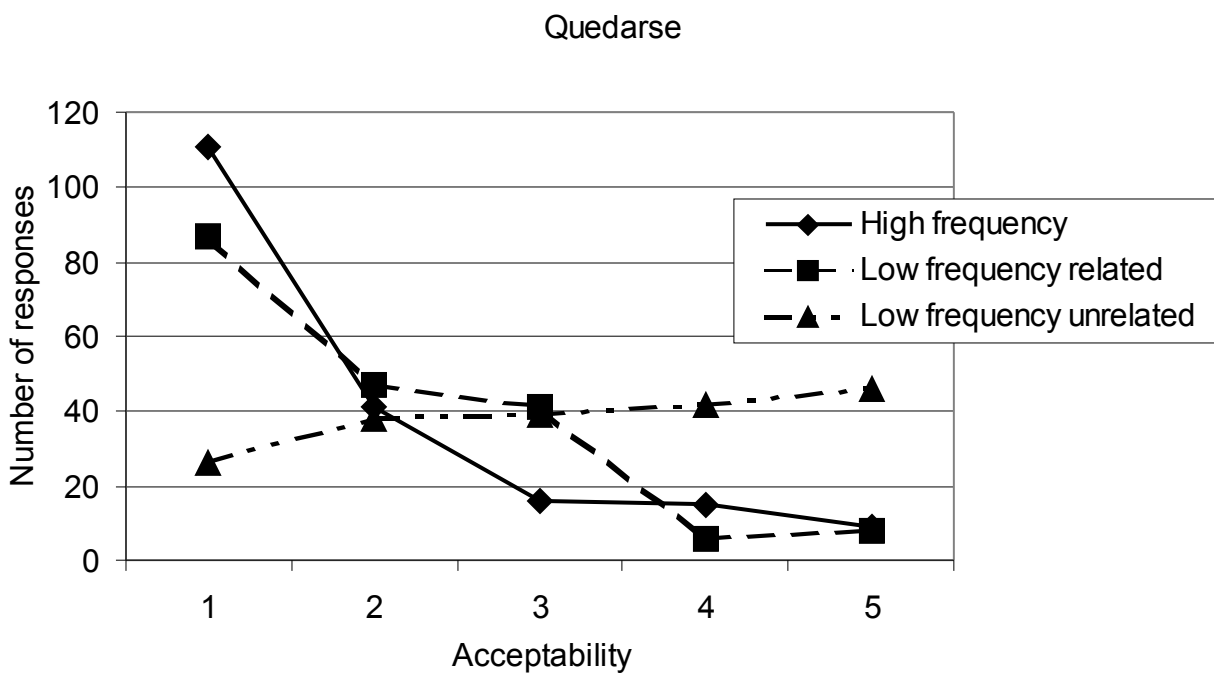


Figure 9

Acceptability ratings for expressions with *quedarse*.

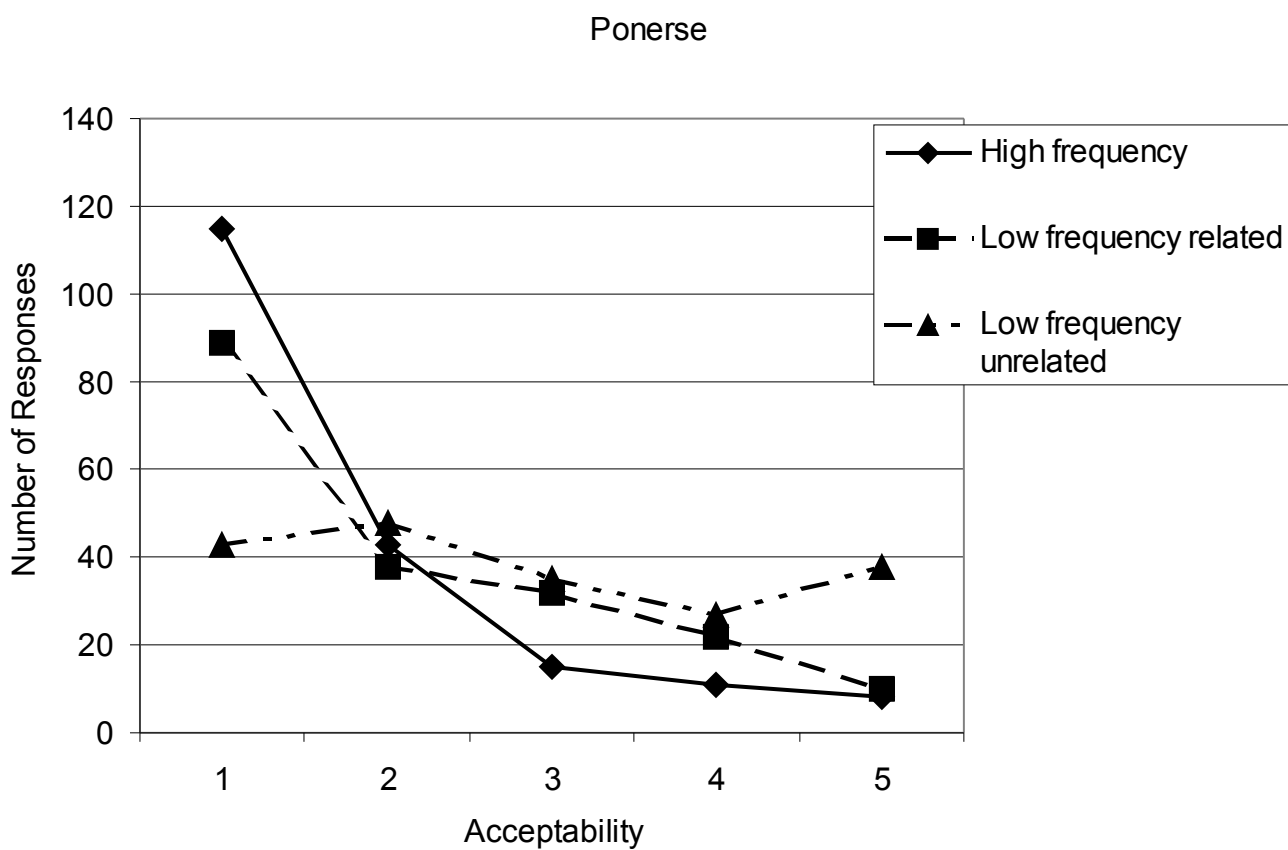


Figure 10

Acceptability ratings for expressions with *ponerse*.