B O R E A L I S
An International Journal of Hispanic Linguistics

Spanish verbs of Becoming and Cluster Concepts

Carlos Benavides
University of Massachusetts Dartmouth

Abstract. The behavior of Spanish verbs of becoming (e.g., *volverse rico* 'to become rich'), also known as pseudocopular or semicopular verbs, has received relatively little attention in the literature. Bybee & Eddington (2006) (B&E) analyze these verbs employing an exemplar model based on frequency and semantic similarity. The current paper presents an analysis of verbs of becoming based on the notion of *cluster concepts* (Jackendoff 2002, 2019, 1983), set within a scale, as an alternative account to B&E's, that explains the behavior of this class of verbs in a more precise and comprehensive way. The current study shows that there are strong tendencies in the use of adjectives with a given verb of becoming, which correlate to their use with the copular verbs *ser/estar*. Crucially, the cluster concept model accounts for uses of adjectives that do not fall into any particular grouping in B&E's model. A corpus study supports the analysis. The current study represents an original contribution in that it is the only study that treats verbs of becoming using the notion of cluster concepts.

Keywords. change of state verbs, cluster concepts, exemplar, copular verbs, similarity

Resumen. El análisis del comportamiento de los verbos de cambio de estado en español (por ejemplo, *volverse rico*), también conocidos como verbos pseudocopulativos o semicopulativos, ha recibido relativamente poca atención en estudios anteriores. Bybee & Eddington (2006) (B&E) analizan estos verbos empleando un modelo basado en ejemplares, el cual a su vez está basado en la frecuencia y la similitud semántica. Este artículo presenta un análisis de los verbos de cambio de estado basado en la noción de *conceptos en cúmulo* (Jackendoff 2002, 2019, 1983), situados dentro de una escala, como una explicación alternativa a la de B&E, que explica el comportamiento de esta clase de verbos de una manera más amplia y precisa. El presente estudio muestra que existen fuertes tendencias en el uso de adjetivos con un verbo de cambio de estado determinado, que se correlacionan con su uso con los verbos copulativos *ser/estar*. Un punto clave es que el modelo de conceptos en cúmulo da cuenta de usos de adjetivos que no entran en ninguna agrupación particular en el modelo de B&E. Un estudio de corpus sirve para respaldar el análisis. El presente trabajo representa una contribución original en la medida en que es el único estudio que trata los verbos de cambio de estado utilizando la noción de conceptos en cúmulo.

Palabras clave. verbos de cambio de estado, conceptos en cúmulo, ejemplar, verbos copulativos, similitud

1. Introduction

This investigation focuses on Spanish verbs of becoming (e.g., *volverse rico* 'to become rich'), a topic which has received relatively little attention in the literature. In one of the most important studies on this topic, Bybee & Eddington (2006) (henceforth B&E) analyze these verbs employing a usage-based model, where they present exemplar-based representations of constructions with verbs of becoming. In the model, analogy to these representations, and not rules that refer to abstract features (e.g., "is permanent," "is essential"), accounts for productive use. Their analysis is based on frequency and semantic similarity, which leads to the organization of dense clusters of

© Carlos Benavides. *Borealis: An International Journal of Hispanic Linguistics*, 2023, 12 / 2. pp. 255-285. https://doi.org/10.7557/1.12.2.7207

semantically related adjectives centered on a high-frequency exemplar. B&E argue that the results of their study support exemplar representations, which are heavily based on usage experience. In particular, they eschew the use of necessary and sufficient conditions or categorical (all-or-none) features as criteria for choosing a particular verb for a particular situation. Thus, B&E discard explanations of use such as, "*ponerse* requires subject agentivity" or "some verbs of becoming express changes that are unexpected or unusual."

The current paper presents an analysis of verbs of becoming based on the notion of *cluster concepts* (Jackendoff 2002, 2019, 1983) as an alternative account to B&E's. The aim of the paper is to show that the notion of cluster concepts, set within a scale, explains the behavior of this class of verbs in a more precise and comprehensive way than B&E's model. The present study shows that there are strong tendencies in the use of adjectives with a given verb of becoming, which correlate to their use with the copular verbs *ser/estar*. The model based on cluster concepts accounts for the selection of adjectives by each verb of becoming (e.g., *nervioso* in *ponerse nervioso* 'become nervous') with a high degree of consistency; has predictive power; dispenses with necessary and sufficient conditions and all-or-none characterizations; and, crucially, the cluster concept model accounts for a significant number of uses of adjectives that do not fall into any particular grouping in B&E's exemplar model. Thus, unlike B&E, the cluster concept model can account for the vast majority of uses of adjectives in the constructions under investigation.

Cluster concepts, also known as cluster conditions, preference rules, preference conditions or default conditions (Jackendoff 2002, 2019, 1983), occur when, given a set of conditions, examples that maximally satisfy the conditions are regarded as more typical than examples that satisfy fewer of the conditions (Jackendoff 2002, 2019). For example, *climbing* involves two independent conditions: (a) an individual is traveling upward, and (b) the individual is moving with effortful grasping motions (clambering). On the most likely interpretation (*Bill climbed (up) the mountain*), both conditions are met. On the other hand, the sentence *Bill climbed down the mountain* violates the first condition and, since snakes cannot clamber, *The snake climbed (up) the tree* violates the second condition. However, both examples are acceptable instances of climbing since each one of them fulfills at least one condition. Nonetheless, if both conditions are violated, as in **The snake climbed down the tree*, the action cannot be characterized as climbing.

Thus, neither of the two conditions is necessary but either is sufficient. According to Jackendoff (2002, 2019, 1983), the key idea is that the default interpretation (*Bill climbed (up) the mountain*), in which both conditions are satisfied, is judged to be more prototypical climbing, while the other acceptable sentences are judged somewhat more marginal but still legitimate sentences. Thus, the more conditions are satisfied, the more likely is the prototypical interpretation. The name *cluster concepts* comes from the idea that conditions cluster together to determine the most prototypical reading. A more detailed explanation of the notion of cluster concepts is presented in § 3.

A corpus study based on the Corpus del Español (CDE, Davies 2002-) was conducted as part of the present study to support the analysis based on cluster concepts. As in B&E, the four Spanish verbs of becoming that occur most frequently with adjectives are examined: *quedarse*, *ponerse*, *volverse*, and *hacerse*. The current study represents an original contribution in that, to the best of our knowledge, it is the only study that treats verbs of becoming using the notion of cluster concepts.

It is important to point out that while both B&E's model and the model developed in this paper deal with the same type of construction ([verb of becoming + adjective]), the application of the cluster conditions is rule-like, whereas in B&E's model, the use of the constructions is determined by similarity to an exemplar. This difference and its relevance are explained in more detail in § 4.

The paper is organized as follows: § 2 reviews B&E's exemplar-based analysis of verbs of becoming. In § 3, the notion of cluster concepts is explained, and § 4 shows how cluster concepts apply to the Spanish verbs of becoming under study, supported by the results of the corpus study. In this section, a comparison is also made between B&E's exemplar model and the model based on cluster concepts. The current study shows that cluster concepts, set on a scale, provide a consistent and predictive explanation for the use of adjectives that occur with verbs of becoming. § 5 discusses future lines of research.

2. The Exemplar Model

This section presents a review of B&E's analysis of verbs of becoming within the exemplar model, as a basis for comparison with the model proposed in the current study, based on cluster concepts. Brief comparisons are made between B&E's exemplar model and the current model in this section, and a full comparison between the two models is presented in § 4. The question may arise as to why this article focuses on B&E's study. This is because, to the best of our knowledge, no other studies have treated verbs of becoming using clusters based on exemplars. The presentation by B&E of those clusters in visual form is a good point of departure for a comparison with cluster concepts, which are also presented visually in this study in the form of a scale. It is very important to emphasize that while B&E present clusters (or groupings) of semantically related adjectives as part of their analysis, the notion of *cluster concepts* is adapted from Jackendoff (2002, 2019, 1983) and is used in the current study to compare it with B&E's use of clusters of adjectives. The innovative contribution of the current study is its treatment of verbs of becoming using the notion of cluster concepts developed by Jackendoff (2002, 2019, 1983).

B&E's study is based on a spoken corpus and a written corpus. The first is a corpus of 1.1 million words of spoken Spanish that consists of transcribed speech from Spain (Marcos Marin 1992) and includes a wide variety of types of oral speech. The written corpus consists of a compilation of fifteen novels (990,000 words) published in the last half of the twentieth century from various Spanish-speaking countries or regions. Thus, the total number of words in the corpora used by B&E is 2,090,000.

In their study, B&E examined 161 adjectives and prepositional phrases (such as *a gusto* 'satisfied') (for a total of 423 tokens) that occur in constructions with the four target verbs of becoming. B&E examined only constructions that occur with an animate subject. As noted above, B&E analyze these verbs employing a usage-based model where they present exemplar representations of periphrastic constructions with verbs of becoming ([verb + adjective]) sequences, such as *ponerse pálido* 'become pale), where analogy to these representations, and not rules that refer to abstract features (e.g.,"is permanent," "is essential"), accounts for productive use. Their analysis is based on frequency and semantic similarity (based on features), which leads to the organization of dense clusters of semantically related adjectives centered on a high-frequency exemplar.

B&E developed their exemplar model after reviewing research on proposed criteria for distinguishing among change-of-state verbs based on the type of change they describe. B&E observe, for example, that certain verbs are thought to express changes that are unexpected or unusual (e.g., *Se quedó asombrada* 'She was amazed') while others convey normal or expected changes (e.g., *Se puso viejo* 'He became old'). Other proposed factors concern whether the change involves an essential or nonessential property (*Se volvió egoísta* 'He/she became selfish') or whether the change results in a temporary or permanent state (*Se volvió enfermizo* 'He became

sickly'). B&E note, however, that the particular features proposed, such as the degree of expectedness, essentiality, and permanence, are not easily defined.

B&E eschew the use of necessary and sufficient conditions or categorical features (such as the ones mentioned above) as criteria for choosing a particular verb for a particular situation. They do so based on arguments such as the following: The original meaning of *quedar* is 'to remain or stay, to be left or left over,' a fairly passive sense, while *poner* is more active, meaning 'put or place'. Thus, *ponerse* would literally mean 'to place oneself,' a meaning that indicates a more active role for the subject. B&E point out that the feature of subject agentivity evident here is neither a necessary nor a sufficient condition for the use of *ponerse*, because there are many instances where no agentivity is involved (e.g., *ponerse nervioso* 'become nervous,' *pálido* 'pale,' *rojo* 'red,' *mal* 'sick,' or *ciego* 'blind'). And some examples with *quedarse* seem parallel to examples with *ponerse*, so the supposed distinction in meaning between the two verbs does not always emerge (e.g., *ponerse triste* and *quedarse triste* have virtually the same meaning).

As in B&E, the present paper does not discuss in detail the background of the verbs of becoming under study, which are also known as pseudocopular, semicopular or pseudoattributive verbs of change (cf. Delbecque & Van Gorp (2015), Morimoto & Pavón Lucero (2004)). The focus is on how cluster concepts can account for the data better than B&E's exemplar model. However, it is important to note that studies that deal with pseudocopular verbs of change (Delbecque & Van Gorp (2015), Van Gorp (2015), Morimoto & Pavón Lucero (2004)) define the behavior of these verbs using criteria that provide a fine-grained meaning of the verbs and, importantly, are similar to the categorical conditions eschewed by B&E. For example, Morimoto & Pavón Lucero (2004) distinguish two different constructions with *quedarse*, one which expresses a punctual inchoative event, and another which expresses a punctual event that delimits the end of a gradual change of state. For its part, the verb *ponerse* denotes a punctual or gradual inchoative event. Thus, by employing what seem to be categorical criteria, these analyses are subject to the same criticism of inadequacy from B&E. As mentioned in § 1, this pitfall is avoided in the present study with the use of cluster concepts. See § 4 for a discussion of the relationship between these semantic criteria and the cluster conditions.

In order to investigate the role semantic features play in determining the occurrence of the verbs of becoming under study, Eddington (1999) applied the following four of the more concrete factors suggested in the literature to characterize verbs of becoming to 1,283 tokens of change-of-state verbs that occurred in a large corpus: (a) whether the verb's predicate is nominal or adjectival; (b) whether the change requires more than twenty-four hours to take place; (c) whether the change occurs with the active participation of the animate experiencer; and (d) whether the noun or adjective of the predicate would be expressed with the copular verbs *ser* or *estar*. A number of tendencies were found. For example, *llegar a ser* is generally used with changes that require more than twenty-four 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, furthermore, that there was a significant amount of overlap; that is, occurrences of these verbs are not necessarily mutually exclusive, and thus, more than one verb may be appropriate in a given context. According to B&E, this overlap is problematic for models based on categorical criteria because, by definition, categories have boundaries, so any overlap indicates that the category boundaries are not well defined.

Eddington (2002) also showed there is overlap among the uses of change-of-state verbs. A questionnaire was given to native Spanish speakers whose task was to indicate which change-of-

state verb they would prefer in a sentence. Results showed that rather than a complementary distribution between the uses of each verb, there was a great deal of overlap in their productive application to new situations. According to B&E, these studies further demonstrate the failure of categorical features to predict the use of change-of-state verbs in Spanish. B&E admit that there are some strong tendencies that allow one to differentiate certain uses of the change-of-state verbs, but hold that the boundaries of the categories of complements to these verbs are fuzzy. I show in § 4 that indeed there are strong tendencies in the use of adjectives with a given verb of becoming, which correlate to their use with *ser/estar*. I show in § 4 as well that these strong tendencies can be formalized through the use of cluster concepts, which allow for flexibility and are not based on categorical, all-or-none judgments. Note that while the correlation between verbs of becoming and *ser/estar* is not new (it was one of the factors studied by Eddington (1999); see Morimoto & Pavón Lucero (2004)), what is new in the current study is that this strong correlation is explained in a principled way through the use of cluster concepts set on a scale, and supported by a systematic corpus study.

Given the results of the studies mentioned above, B&E argue 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. Such an approach would allow for the existence of 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.

B&E's analysis of the four target verbs (*quedarse*, *ponerse*, *volverse*, and *hacerse*) leads to the organization of tokens into the above-mentioned clusters of semantically related adjectives centered on a high-frequency exemplar, but it does so for just two of the verbs (*quedarse* and *ponerse*). By exploring the items that appear in the adjective position in these two constructions, B&E produced a clustering that can presumably predict subsequent uses of the constructions. However, B&E do not show exactly how that prediction is implemented. Presumably, the adjectives clustering around a central member are more likely to be used with a given verb of becoming. Nevertheless, this is not shown or stated explicitly. Crucially, B&E found no exemplar-based clusters for adjectives that occur with *volverse* and *hacerse*, and they also found isolated uses for some adjectives that occur with *quedarse* and *ponerse*. I show in § 4 that the use of what are isolated instances in the B&E model finds an explanation under the cluster concepts model.

Figure 1 shows a cluster, presented by B&E, centering on the adjective *solo* 'alone,' as in *quedarse solo* 'to end up alone'. This is the adjective with the highest frequency, and thus, as the exemplar, it is placed in the center, to indicate its hypothesized status as the central member. According to B&E, high-frequency exemplars could be thought of as prototypical since they exhibit most of the features shared by the other members. 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 B&E use to indicate that it has an opposite value in one feature and constitutes a related concept.

Figure 1. Cluster centering on quedarse solo [B&E, p. 332]

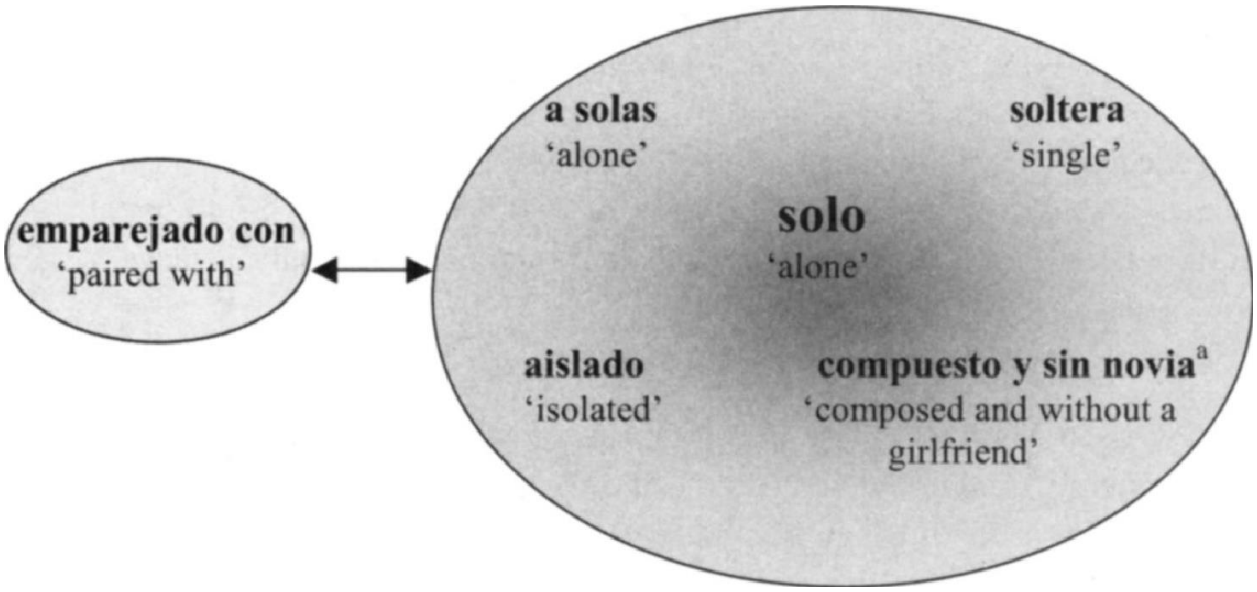
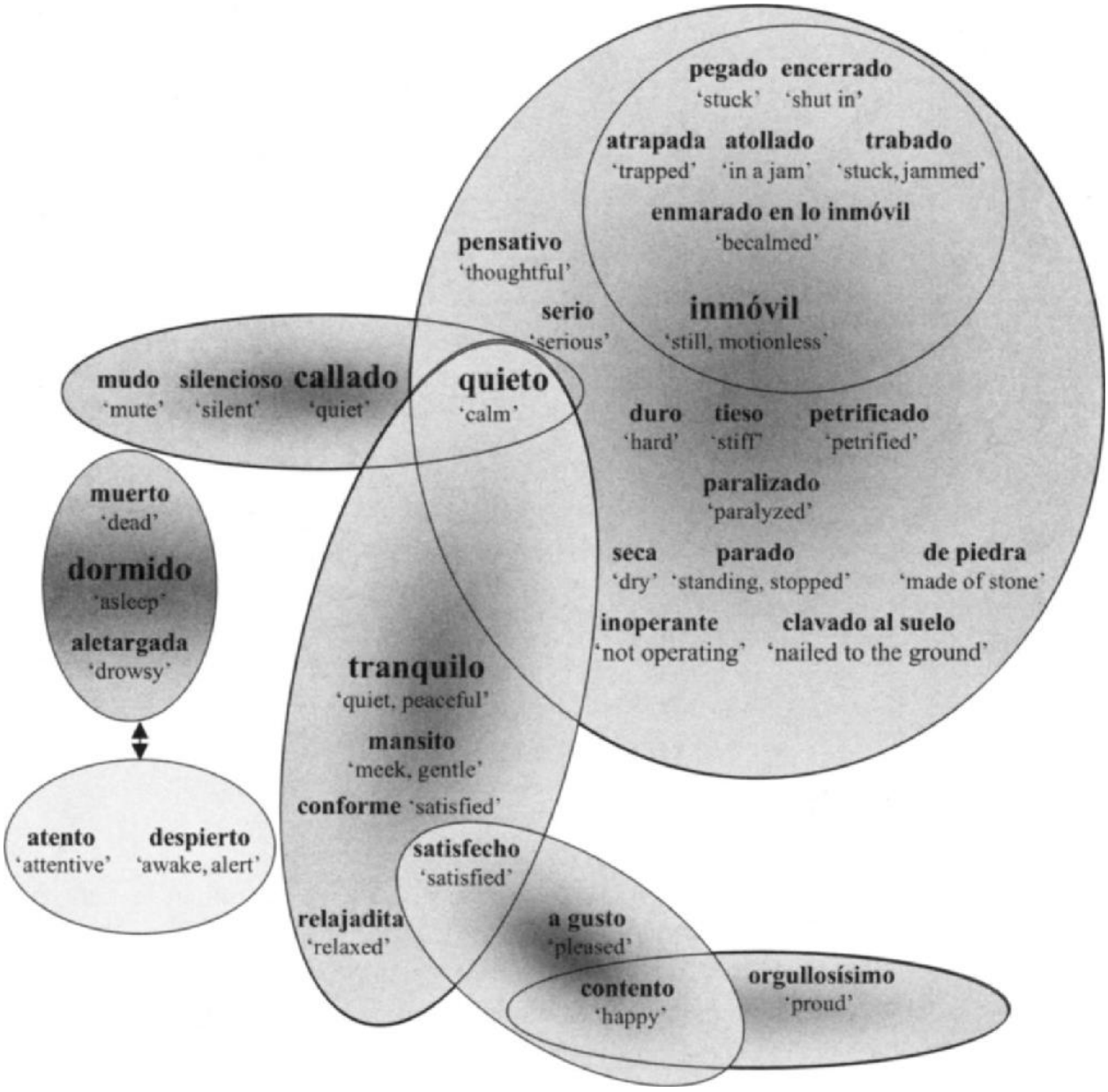


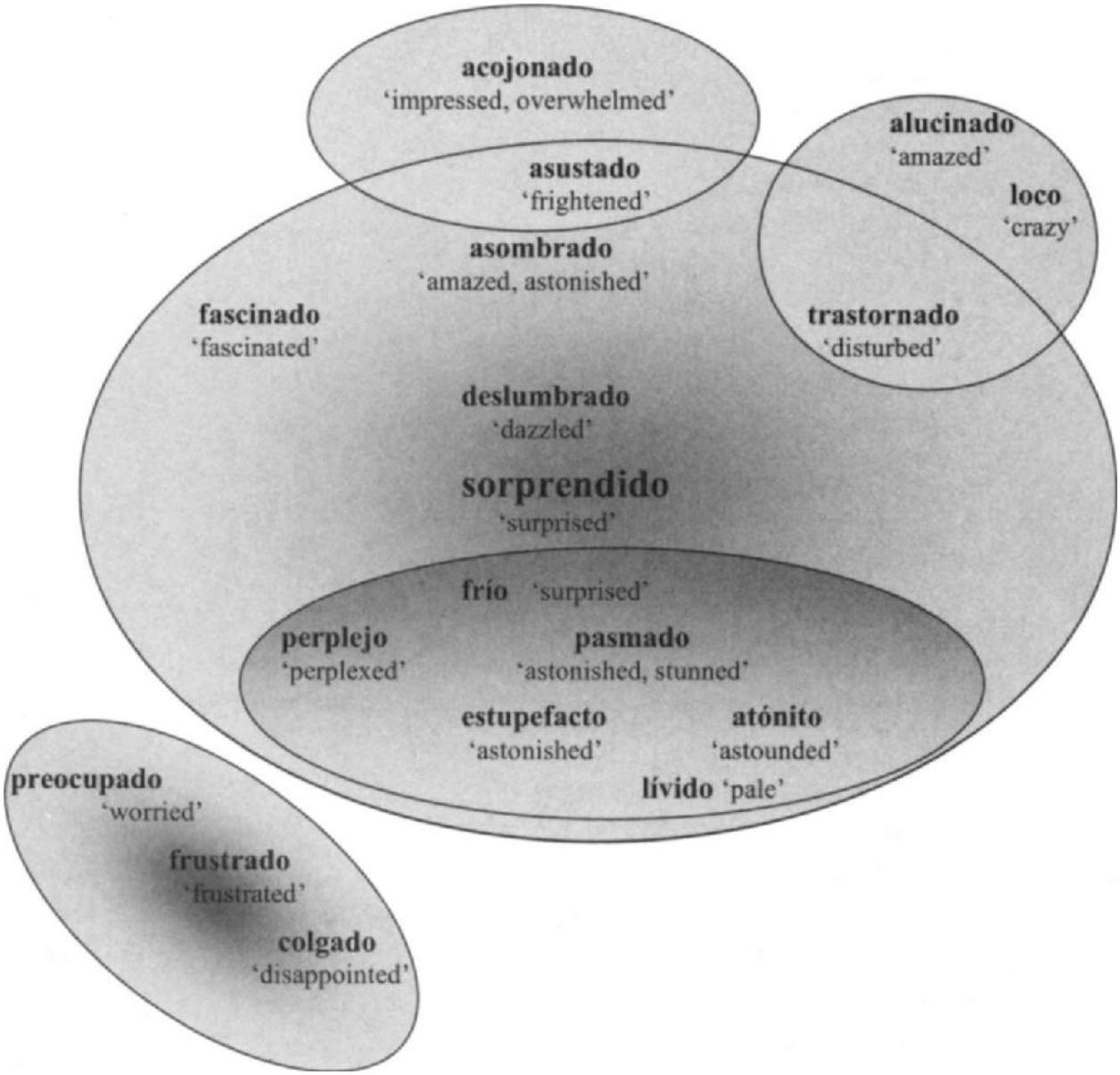
Figure 2 shows one of the most complex clusters posited by B&E, centering on *quieto* 'calm' (as in *quedarse quieto* 'to become calm'); it is the cluster that encompasses the most types and tokens. This cluster is so complex because there are five sub-clusters that overlap and are related. The figure shows the most frequent members of these five clusters. In addition to frequency, q*uieto* was chosen by B&E as the central member because it shares features with the other major clusters for *callado* 'quiet,' *inmóvil* 'still, motionless,' and *tranquilo* 'calm, peaceful'.

Figure 2. Cluster centering on quedarse quieto **[***B&E, p. 334]*



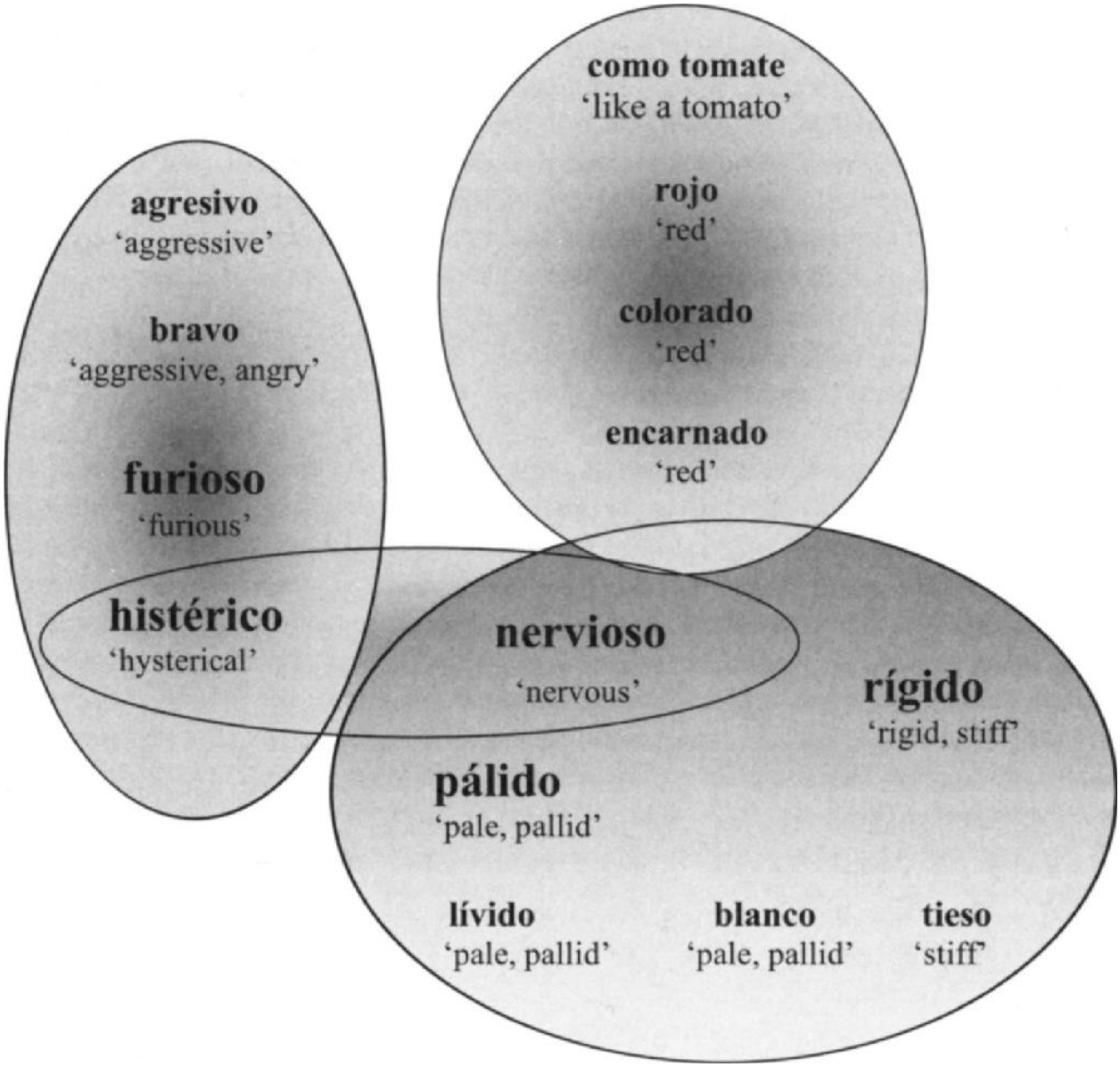
Note that the adjectives in the two clusters above have in common an absence of movement, agitation or company. However, there is a cluster in B&E's study which does not share features with any of the other clusters with *quedarse* (nor is it the opposite of any other cluster). This is the cluster shown in Figure 3, which centers on *sorprendido* 'surprised'. Thus, say B&E, it appears that the various categories of adjectives used with *quedarse* need not be related to one another. However, this means that in the exemplar-based model there are not only isolated instances of adjectives, but also isolated clusters of adjectives. In contrast, as seen in § 4, in the model proposed in this study, most adjectives used with *quedarse* are shown to have something in common.

Figure 3. Cluster centering on quedarse sorprendido [B&E, p. 337]



Finally, Figure 4 shows the cluster centering on *ponerse nervioso* 'to become nervous,' which follows the pattern for the previous clusters. B&E presented a few more clusters with *quedarse* and *ponerse*, but the ones shown here are sufficient to grasp the general pattern.

Figure 4. Cluster centering on ponerse nervioso [B&E, p. 340]



Note that the cluster Figure 4 does not capture the fact that *agresivo* 'aggressive' is used frequently not only with *ponerse* but also with *volverse* (see § 4). This may have to do in part with the fact that there was only 1 token for *agresivo* in B&E's corpus, compared to the 463 tokens found in the Corpus del Español (see more details about the corpora in § 4).

As noted above, the clusters presented by B&E occurred just with *quedarse* and *ponerse*, not with *volverse* and *hacerse*. As indicated by B&E, this occurred for a particular reason: lack of similarity to other adjectives. The verbs *volverse* and *hacerse* are used with more diverse sets of adjectives, and B&E found that these adjectives do not fall into clusters. That is, the adjectives that combine with *volverse* and *hacerse* (e.g., *volverse esquivo* 'to become evasive,' *hacerse responsable* 'to become responsible'), as well as some adjectives with *quedarse* and *ponerse* (e.g.,

embarazada ‘pregnant,’ *pesado* ‘annoying’), are semantically isolated uses that do not bear a semantic resemblance to any other exemplars of a given construction. Thus, they are not accounted for in the exemplar model. In contrast, as seen in § 4, the current model accounts for the behavior of adjectives, including the “isolated” uses, in all four constructions, on the basis of their position on a scale based on cluster concepts, and their association in meaning with *ser* and *estar* (see examples in § 4).

Second, a group of adjectives describing decidedly negative personal behavior has as its most frequent type *pesado* 'annoying,' as in *ponerse pesado* 'to become annoying.' According to B&E, several groupings for these adjectives (e.g., *bobo* ‘silly,’ *insolente* ‘insolent, rude’) could be proposed, some overlapping slightly and some not. However, B&E note that the exact structuring of this category is less apparent than for some of the others with *ponerse*, and that the relation of this group with other adjectives with *ponerse* seems to be somewhat arbitrary, leaving the issue for future research.

Note that while the number of isolated uses in B&E’s study was small, what is crucial is that no clusters at all could be formed with the verbs *volverse* and *hacerse*. Considering that a large number of adjectives occur with these two verbs, this represents a significant gap in the explanatory range of the exemplar-based model. In contrast, as shown in § 4, there are no gaps in the explanation offered by the cluster concepts model, as it accounts for the vast majority of adjectives used with all four verbs of becoming under study.

Finally, B&E also conducted a multidimensional analysis, which involves the mapping of a group of adjectives in terms of their semantic similarity. The groupings emerging from the analysis show the adjective *sorprendido* ‘surprised’ as central to a group of adjectives, much as predicted. The adjective *inmóvil* ‘immobile’ however, which was also predicted by B&E to occupy a central position in its category, is located in a relatively marginal position, in a similar way to the isolated cases discussed above.

3. Cluster Concepts

As noted in the § 1, the key idea behind cluster concepts is that the more conditions for the definition of a word (e.g., the verb *climb*) are satisfied, the more likely is the prototypical interpretation. Thus, conditions cluster together to determine the most prototypical reading. As Jackendoff (2002) notes, the notion of cluster concepts can also be applied to nouns and prepositions. For instance, a prototypical chair satisfies the condition of a prototypical form and a second condition of a standard function. Objects that have the proper function but the wrong form, such as beanbag chairs, as well as objects that have the right form but cannot fulfill the function, say chairs made of newspaper, are both more marginal instances of the category. An object that violates both conditions, say a pile of crumpled newspaper, is definitely not a chair. To take an example involving three conditions, the concept *mother* includes the woman who contributes genetic material, the woman who bears the child, and the woman who raises the child. As noted by Jackendoff (2002), in today’s world of genetic engineering and adoption (and I would add surrogate mothers), not all three of these conditions always coincide to define *mother*, and so the term is not always used prototypically. That is, instances of mothers that satisfy fewer than the three conditions will be considered less typical instances.

Now consider the preposition *in*, which has a geometric and a functional component. The geometrical condition stipulates that if an object X is in a place or object Y, X must be geometrically within the interior of Y. The functional condition, on the other hand, can be characterized as “containment”: roughly, X is not attached to Y, but if one moves Y, X is physically

forced to move with it. The prototypical instances of *in*, say a pebble in a bottle, satisfy both conditions. However, a pear in a bowl that is on top of other fruits, and a knife stuck in a (large) piece of cheese, while satisfying the functional condition, violate the geometrical condition, as they are not fully within the object. The functional condition is violated by examples such as a circle drawn on top of a square and overlapping with it, with neither figure being completely inside of the other. In any case, only instances that satisfy both conditions are considered the typical ones.

As noted above, cluster concepts are also known as *preference rules* or *preference conditions*. The notion of *preference* can be illustrated with objects ranging from cups to bowls, which can differ in height, width, and shape. The property of having a handle can be considered a preference rule so that, for example, an object that because of its height and width is more like a bowl, may instead be judged as a cup when it is given a handle (cf. ten Hacken 2014, Jackendoff 1983). That is, having a handle prompts a preference to consider the object more like a cup than a bowl. Thus, preference rules help determine the categorization of objects. The fact that cluster concepts/conditions are also known as preference *rules* is important. They are rules in that they involve the satisfaction of conditions in order to reach an interpretation. This fact is relevant because it establishes a clear contrast with exemplar-based theories, which are based on similarity to an exemplar rather than rules. This contrast is developed further in § 4, 5.

As seen from the examples above, cluster concepts contribute to explain the variability of word meanings, while keeping them within limits. Crucially, cluster concepts allow for flexibility and are not based on categorical, all-or-none judgments. In the current paper, the notion of cluster concepts is applied to adjectives that occur in constructions with the target verbs of becoming, in order to better understand and predict their behavior.

Jackendoff (2002) notes that the notion of cluster concepts is plausible on broader grounds, citing visual perception, phonetic perception, musical cognition, and Gricean implicature as other areas where cluster concepts apply. Furthermore, I have noticed that the reasoning behind cluster concepts is similar to the one supporting Dowty's (1991) Argument Selection Principle (ASP), whereby the subject of a sentence is the verbal argument with the greatest number of Proto-Agent properties (such as volitional involvement or causation), while the object is the argument with the most Proto-Patient properties (such as undergoing a change of state or being causally affected). Consistent with the principles behind cluster concepts, an argument does not have to satisfy all the properties of a given Proto-Role in order to be assigned that role. This shows that, in one way or another, the underlying notion of cluster concepts has been applied in linguistic analysis for a relatively long time.

4. Cluster Conditions and Spanish Verbs of Becoming

In this section I show how cluster concepts provide a precise, consistent, and predictive explanation for the use of adjectives that occur with verbs of becoming: the more conditions an adjective satisfies, the more likely it is to appear with a given verb. Because the notion of cluster concepts is crucially based on a set of conditions, henceforth the model is also referred to as *cluster conditions*. As in B&E, I analyze constructions with only animate subjects, with the four verbs of becoming that occur most frequently with adjectives: *quedarse*, *ponerse*, *volverse*, and *hacerse*. Verbs that occur mostly with nouns (*convertirse en* 'become,' *transformarse en* 'transform into,' and *llegar a ser* 'come to be') are not examined (cf. Eddington 1999). It is important to point out that the author of the current paper is a native speaker of Spanish. Corpus results, as well as the analysis developed throughout this section, are informed by the author's own intuitions. This is

helpful, for example, when it comes to detecting nuances in the use and meaning of verbs and adjectives found in the corpus, in order to select which are appropriate to be examined.

The cluster conditions proposed for the current model are shown in (1).

(1) Cluster conditions
–gradual change
–long-term change
–sustained action that leads to a change

Elements in these conditions such as "gradual" and "long-term" (or similar terms) are mentioned as relevant semantic factors for verbs of becoming in B&E and other sources. For example, Morimoto & Pavón Lucero (2004) point out that some verbs of becoming indicate a "gradual change of state," while B&E mention that other proposed factors concern whether the change results in a permanent state. As for the notion of "change" itself, this is the key concept that defines these verbs (cf. Morimoto & Pavón Lucero (2004)), so it should be a part of all conditions. These conditions were formulated in their present form because they help make relevant distinctions among all four verbs of becoming, in contrast to other potential cluster conditions which are not helpful to make such distinctions, and may even lead to inaccurate predictions.

For example, the notion of volition/intention (or agency) has not been added as a condition because it does not seem to make a helpful distinction of use among all verbs of becoming. For instance, such a cluster condition might help explain the difference between "*volverse responsable*" (no volition) and "*hacerse responsable*" (volition), but it does not help to distinguish these verbs from the other two verbs of becoming, *quedarse* and *ponerse*. In addition, it can be argued that *quedarse* also involves no volition, while *ponerse* does involve volition (cf. B&E, Van Gorp (2015)). This would lead to the wrong prediction that *quedarse* patterns with *volverse*, while *ponerse* patterns with *hacerse*. As seen below, it is *quedarse*/*ponerse*, on the one hand, and *volverse*/*hacerse*, on the other hand, that seem to pattern together. The notion of volition/intention has a role in determining whether to use one verb of becoming or another, as shown above, but, again, it is not helpful in making a distinction among all four verbs of becoming, which is necessary for the inclusion of a feature in the cluster conditions.

Similarly, Delbecque & Van Gorp (2015) posit the notion of progression as characterizing *hacerse* and regression for *volverse*, but these features do not help either. They help distinguish *volverse* from *hacerse*, but they have nothing to say about the other two verbs, so they do not lead to helpful distinctions among all four verbs. For another example, according to Van Gorp (2015), *ponerse* profiles the transition point and the adoption of the new state, thus yielding a dynamic view of the subject entity entering a new state, while *quedarse* profiles the static outcome of a change event by highlighting only the resultant state of the change event, caused by some separation or loss. Again, however, these criteria do not help make a helpful distinction in the use of all four verbs of becoming. Furthermore, these categorical-like criteria attempt to account for the complementary distribution of these verbs (as noted by Eddington (1999, 2002) and B&E), but are not helpful in explaining the overlap in usage.

The more conditions in (1) are satisfied by an adjective, the more likely the adjective is to be used with *hacerse*, and the fewer conditions are satisfied, the more likely the adjective is to be used with *quedarse*. Thus, *quedarse* and *hacerse* are on opposite ends of the proposed scale, shown in Table 1, with *volverse* and *ponerse* in the middle. Note that it is the verbs that determine the

change of state, which is instantiated in the conditions. Thus, it can be said that the adjectives satisfy conditions imposed by the verbs in the form of cluster concepts. The first two conditions seem to be key to determine use with *hacerse*, and the remaining one (sustained action) seems to be helpful to confirm, strengthen or reinforce the classification. For example, once a speaker has determined that a given adjective satisfies the first two conditions (gradual change and long-term change), the speaker knows that the adjective is very likely to occur with *hacerse*. On the other hand, if the adjective does not satisfy any of the conditions, it is used with *quedarse*.

Table 1. Scale formed by cluster conditions

	1	2	3	4
Condition/verb	quedarse atónito 'become astounded, astonished'	ponerse enfermo 'become sick'	volverse agresivo 'become aggressive'	hacerse responsable 'become responsible'
gradual change	0	<	<	+
long-term change	0	<	<	+
sustained action	0	0	<	+

This is shown in Table 1, which illustrates how the cluster conditions can generate a scale where, from left to right, adjectives satisfy more and more conditions. Thus, on one end, s*e hizo responsable* 'he became responsible,' satisfies all three conditions, while on the opposite end, *se quedó atónito* 'he was astounded' does not satisfy any of them. Note that each of the verbs on the scale appears with a representative adjective. The symbols make reference to the adjectives that occur with these verbs, and indicate the following: "+" = the adjective satisfies the condition to the highest degree; "<" = satisfies the condition to a higher degree than adjectives used with the verb on the left; and "0" = does not satisfy the condition. Thus, *hacerse*, which has a "+" for all three conditions, satisfies the conditions to the highest degree, while *quedarse*, on the other end of the spectrum, has a "0" for all conditions and thus can be said to satisfy the conditions to the lowest degree.

The scale in Table 1 also accounts for overlap in the selection of adjectives (cf. B&E, Eddington 1999, 2002), and shows that the overlap is limited and systematic: there is overlap between adjectives that fall nearby in the scale, such as those used with both *ponerse* and *volverse* or with both *quedarse* and *ponerse*, but there is hardly any overlap when adjectives are on opposite ends of the scale (those used with both *quedarse* and *hacerse*). This happens because adjectives that are closer to each other on the scale are also closer in the number of conditions they satisfy. For example, *arrogante* 'arrogant' is commonly used with both *volverse* and *hacerse* but is rarely used with *quedarse* or *ponerse* because it satisfies at least two conditions in a clear way (gradual change and long-term change), and possibly the third one as well, in certain situations. Conversely, *contento* 'happy' is frequently used with both *quedarse* and *ponerse* but rarely with *volverse* or

hacerse because it does not satisfy any of the conditions. Corpus data presented below support this analysis.

It is important to emphasize that the scale in Table 1 reflects tendencies in the usage of the adjectives with verbs of becoming and, unlike categorical criteria, does not deal in absolutes. As noted in § 2, the strong tendencies in the use of adjectives with a given verb of becoming, which correlate to their use with *ser/estar*, are formalized through the use of the scale in Table 1, which allows for flexibility and is not based on categorical, all-or-none judgments. This is a key advantage of a continuum over all-or-none criteria. Thus, for example, while *arrogante* 'arrogant' is used more frequently with both *volverse* and *hacerse* (as corpus results show), for some native speakers its use with *ponerse* may sound natural.

Note that *quedarse* does not always have a change-of-state meaning; it can have a stative meaning when used with practically any of the adjectives it combines with, depending on the context. For example, *Ella se quedó atónita al escuchar la noticia* 'She was astounded upon hearing the news' has a meaning that involves change, while *Ella se quedó atónita durante todo el espectáculo* 'She remained astounded during the entire show' has a stative meaning. There is also a stative use of *quedarse* in utterances such as *Me quedé despierta toda la noche* 'I stayed up all night.' However, the stative use of *quedarse* is rare with most of the adjectives examined here. Returning to the adjective *contento*, while *quedarse contento* sometimes expresses permanence in a state—a state that has possibly, but not necessarily, been the consequence of a change—it does express a change of state in many of its uses. Taking an example from the corpus, in *Le dije que sí y ella se quedó contenta* 'I told her "yes" and she got happy,' the person becomes happy due to what was told to her, and thus a change of state is expressed.

Another way of grasping how verbs satisfy the conditions in Table 1 is to posit a scale for each condition individually. Thus, for example, the scale for "gradual change" would be as in (2), with *quedarse* and *ponerse* at the low end of the scale and *volverse* and *hacerse* at the high end. The advantage of the full scale in Table 1 is that it captures the application of all conditions at the same time in a succinct way.

(2)Scale for one condition

[- gradual change]			[+ gradual change]
I----------I			
quedarse	ponerse	volverse	hacerse

4.1. The Interaction Between Ser and Estar and Spanish Verbs of Becoming

4.1.1. Corpus Study

In order to explore the interaction between the copular verbs *ser* and *estar* and Spanish verbs of becoming, a corpus study based on the Corpus del Español (CDE, Davies 2002-) was conducted as part of the present study to support the analysis based on cluster concepts. The CDE consists of two corpora: an online corpus that consists of more than 100 million words in more than 20,000 Spanish texts from the 1200's to the 1900's in four registers (oral, news, fiction, and academic), labeled "Genre/Historical"; and a web-based corpus, labeled "Web/Dialects," containing 2 billion words of Spanish taken from 2 million web pages from 21 different Spanish-speaking countries, from the past 3-4 years. Only data from the latter corpus were used for the current study. While

the Genre/Historical corpus has an oral component, which the Web/Dialects corpus lacks, I decided to conduct the study using the latter due to the large difference in size between corpora; the Web/Dialects corpus is 20 times as large as the Genre/Historical corpus. The size difference between the Web/Dialects corpus and the corpus used by B&E (2,090,000 words) is also noteworthy. The former is 957 times larger than the latter.

In the current study, 42 adjectives were examined, many of which coincide with adjectives examined in B&E. This number is relatively low compared to the number of items examined by B&E, due to the exclusion of several types of items from their study because of difficulties in their examination. The following are the types of items in B&E's study that were excluded from the current study: 1) Items that contain a prepositional phrase following the verb (e.g., *a solas* in *quedarse a solas* 'come to be alone') rather than an adjective. 2) Adjectives with an extremely low frequency in both B&E's and the current study's corpora (e.g., *lívido* 'pale,' *pasmado* 'stunned'). Note that a considerable number of adjectives in B&E's study had a frequency of 1 token (*serio* 'serious,' *calvo* 'bald'). In addition, some of these low frequency adjectives have obscure or unclear meanings (e.g., *gótico* 'Gothic,' *borde* 'nasty'), which makes it hard to make reliable generalizations about them. 3) Polysemous adjectives whose senses usually occur with different verbs of becoming because they clearly have different meanings. For example, *serio* can mean both 'with a serious facial expression' as in *Se pusieron serios cuando entró el presidente* 'They became serious when the president walked in,' and 'responsible,' as in *Ella se volvió una persona muy seria en sus negocios*, 'She became a very responsible person in her business dealings.' In its former use, *serio* usually occurs with *quedarse/ponerse*, and in its latter use, with *volverse/hacerse*. The same goes for, say, *orgulloso*, which can mean 'proud' when used with *quedarse/ponerse*, but usually means 'conceited' when used with *volverse/hacerse*.

This situation could potentially be a confounding factor, especially when examining thousands of tokens from a large corpus. This problem probably did not arise in B&E's study due to the small size of the corpus. However, given the large size of the Web/Dialects corpus, it was decided to examine only adjectives with a single, clear meaning. Nevertheless, adjectives such as *serio* are interesting and they would be worth examining in future studies. 4) Adjectives which in the Web/Dialects corpus occur largely with inanimate or abstract subjects when used with a verb of becoming (e.g., *estéril* in *la teoría se vuelve estéril* 'the theory becomes sterile'). Recall that, as in B&E, only constructions that occur with an animate subject were examined in the current study. 5) Adjectives, such as *soltero* 'single,' that express a permanence in a state rather than a change of state, in any reading. Thus, *Se quedó soltero* 'He remained single' is a use of stative *quedarse* that implies that the expected situation would have been the opposite (having married). On the other hand, some adjectives were added that were not examined in B&E, that occur in constructions with clearly animate subjects (e.g., *conformista* 'conformist,' *enfermizo* 'sickly').

While it would be ideal to make a comparison with each and every one of the adjectives in B&E's study, the reasons for excluding adjectives from their study mentioned above, make it quite difficult to make such a comparison. While the total number of adjectives (types) examined in the current study was relatively low compared to B&E's, note the enormous difference in size between both the corpora (957 to 1) and the number of tokens studied. While B&E examined 423 tokens, the current study examined 40,926 tokens of adjectives occurring with any of the target verbs, a proportion of 97 to 1. (The number of tokens of the use of the adjectives with *ser/estar* yielded by the corpus was even higher, 210,145). Recall that B&E also included prepositional phrases in their analysis. This means that the difference between studies in number of tokens of adjectives examined is even larger. Note as well that the number of tokens of change-of-state verbs from a

corpus used in Eddington's (1999) study, 1,283 tokens, is also quite small when compared to the number of tokens examined in the current study. These differences in token numbers compensate in a significant way for the lower number of types studied in the present article, and contribute to give validity to the results of this study.

Two separate searches were done in the Web/Dialects corpus for each target adjective: 1) a search for the frequency of occurrence of each adjective with *ser* and *estar*; 2) a search for the frequency of occurrence of each adjective with each of the four verbs of becoming under study. All forms of the adjectives and verbs yielded by the corpus were considered, including feminine and masculine, singular and plural, as well as all tenses, aspectual variants and moods (e.g., *se pone nerviosa*, *estaba nervioso*, *se quedarían nerviosos*, *se ponga nerviosa*). The searches were done only with the pronoun *se* (and not, for example, with the pronouns *me* or *nos*). Only adjectives with a total frequency of at least 90 tokens occurring with *ser*/*estar* in the corpus were examined. The results of the corpus searches are discussed in the next section.

4.1.2. Corpus Study Results

Even though B&E argue that the use of the copular verbs *ser* or *estar* with a given adjective cannot be used as a reliable criterion to determine what verb of becoming that adjective occurs with, the corpus study results suggest otherwise: there seems to be a strong correlation between the use of an adjective with *ser* or *estar* and the verbs of becoming the adjective occurs with. As seen in Tables 2 and 3, which present the numerical results of both searches, adjectives that are used mostly with *estar* (e.g., *está perplejo* 'is perplexed') tend to occur more frequently with *quedarse*/*ponerse*, while adjectives that are used mostly with *ser* (e.g., *es arrogante* 'arrogant') tend to occur with *volverse/hacerse* more commonly. (Note in Tables 2, 3 that the highest number of occurrences for each adjective with a given verb (*ser* or *estar*) has been set in boldface. For most adjectives, frequency of occurrence with a given verb has also been highlighted.)

For example, *furioso* 'furious' occurs mostly with *estar* and, as expected, it occurs more frequently with *quedarse* and *ponerse*, and *asombrado* 'amazed' occurs overwhelmingly with both *estar* and *quedarse*. On the other hand, *adicto* 'addicted' occurs overwhelmingly with *ser* and also with both *volverse* and *hacerse*. Some adjectives occur mostly with *ser* or *estar*, and do not occur at all with their non-corresponding verbs, *quedarse/ponerse* on the one hand, and *volverse/hacerse* on the other hand. For example, *conformista* 'conformist' and *esquivo* 'evasive' occur mostly with *ser*, and do not occur at all with *quedarse/ponerse*, while *estupefacto* 'stupefied,' and *perplejo* 'perplexed' occur mostly with *estar*, and do not have any occurrences with *volverse/hacerse*.

Note that a sentence such as *A veces te pones de un conformista que me saca de quicio* 'You sometimes become into such a conformist that you drive me crazy,' seems to involve a special, conventionalized structure. We do not see this type of structure with similar adjectives that, like *conformista*, do not usually occur with *ponerse*; for example, **A veces te pones de un adicto/digno/responsable que me sacas de quicio* 'You sometimes become into such an addicted/honorable/responsible (person) that you drive me crazy.' Thus, this type of construction, which could occur with some or all of the three remaining verbs of becoming, seems to be infrequent and idiosyncratic.

There are also cases in between the extremes mentioned above. For instance, *arrogante* 'arrogant' and *enfermizo* 'sickly' occur overwhelmingly with *ser*, though not as frequently with *volverse/hacerse.* Still, the trend whereby adjectives that occur mostly with *ser/estar* also tend to occur with a corresponding verb of becoming, is followed. Overall, the trend is predictive of the use of a given adjective with a given verb of becoming.

The "middle ground" adjectives in Tables 2, 3 are the ones for which there was a predominance of use with *ser* or *estar*, but which either had significant overlap or very few uses with any of the four verbs. For instance, *agresivo* 'aggressive' occurs more with *ser*, and also occurs with both *ponerse* and *volverse*, and *desnutrido* 'malnourished' occurs mostly with *estar*, but does not occur with any of the verbs of becoming. However, these should not be considered exceptions (see below) because they do not clearly pattern with their non-corresponding verbs.

Table 2. Summary of corpus results

Adjectives (42)	% of adjectives
Confirm trend (e.g., adicto, asombrado, arrogante, callado, contento, dormido, embarazada, enfermizo, esquivo, furioso, nervioso)	**79% (33/42)** with *estar*: 22 with *ser*: 11
Middle ground (agresivo, delgado, desnutrido, indignado, insolente, tozudo)	**14% (6/42)**
Possible Exceptions (calvo, loco, terco)	**7% (3/42)**

Table 3. Adjectives and their use with estar, ser, and the four verbs of becoming

	Adjectives	**Used with *estar***	**Used with *ser***
1	adicto 'addicted'	164 quedarse – 1 ponerse – 0	**5,975** **volverse- 419** **hacerse – 249**
2	agresivo 'aggressive'	69 quedarse - 0 **ponerse – 249**	**3,477** **volverse- 201** hacerse – 13
3	alucinado 'amazed'	**184** **quedarse - 64** ponerse – 0	22 volverse – 0 hacerse - 0
4	arrogante 'arrogant'	2 quedarse - 0 ponerse - 2	**1,054** **volverse - 44** **hacerse - 13**
5	asombrado 'amazed'	**894** **quedarse – 354** ponerse – 1	18 (most passive) volverse – 1 hacerse - 1
6	asustado 'frightened'	**2,932** **quedarse - 36** ponerse – 2	25 (most passive) volverse - 2 hacerse - 0
7	atónito 'astounded'	**237** **quedarse - 299** ponerse – 0	5 volverse - 1 hacerse - 0
8	callado 'quiet'	**1,518**	527 (many passive)

		quedarse – 2,508 ponerse – 1	volverse – 6 hacerse - 0
9	calvo ‘bald’	112 total **quedarse - 63** ponerse - 2	**493 total** volverse – 3 hacerse - 0
10	conformista ‘conformist’	1 quedarse - 0 ponerse - 0	**753** **volverse – 23** hacerse - 1
11	contento ‘happy’	**20,567** **quedarse - 149** **ponerse - 471**	108 volverse – 3 hacerse - 0
12	delgado ‘slim’	919 quedarse – 5 **ponerse – 8**	**2,921** **volverse – 18** hacerse – 5
13	deslumbrado ‘dazzled’	**88 total** **quedarse – 16** ponerse - 0	66 (most passive) volverse – 0 hacerse – 0
14	desnutrido ‘malnourished’	**427** quedarse – 0 ponerse - 0	70 volverse - 1 hacerse – 0
15	digno ‘honorable, worthy of’	32 total quedarse – 1 ponerse - 11	**16,405** **volverse – 8** **hacerse - 69**
16	dormido ‘asleep’	**4,928** **quedarse – 2,859** ponerse - 0	85 (many passive) volverse – 0 hacerse - 0
17	egoísta ‘selfish’	3 quedarse – 0 ponerse - 1	**3,853** **volverse – 50** hacerse - 5
18	embarazada ‘pregnant’	**21,473** **quedarse – 906** ponerse - 1	81 (most passive) volverse – 0 hacerse - 0
19	enfermizo ‘sickly’	4 quedarse – 0 ponerse - 0	**385** **volverse – 16** **hacerse - 3**
20	enfermo ‘sick’	**15,551** quedarse – 10 **ponerse - 333**	1,029 volverse – 6 hacerse - 2
21	enojado ‘angry’	**4,192** **quedarse – 9** **ponerse - 15**	47 (some are used with *go* or *leave*, not *ser*; e.g., *se fue enojado*; ‘he left angry’) volverse – 2 hacerse - 0
22	ensimismado ‘introverted’	**92** **quedarse – 16**	22 volverse –3

		ponerse - 0	hacerse - 0
23	esquivo ‘evasive’	10 quedarse – 0 ponerse - 0	**704** **volverse – 11** hacerse **- 7**
24	estupefacto ‘stupefied’	**113** **quedarse – 178** ponerse - 0	2 volverse – 0 hacerse - 0
25	fascinado ‘fascinated’	**1,399** **quedarse – 94** ponerse - 0	58 (most passive) volverse – 0 hacerse - 0
26	furioso ‘furious’	**1,270** quedarse – 13 **ponerse - 628**	82 volverse – 16 hacerse - 2
27	histérico ‘hysterical’	**141** quedarse – 1 **ponerse - 200**	132 volverse - 20 hacerse - 0
28	idiota ‘idiot’	31 quedarse – 1 ponerse - 8	**2,518** **volverse - 22** hacerse - 4
29	indignado ‘indignant’	**2,136** quedarse - 2 ponerse - 0	48 volverse - 3 hacerse - 0
30	insolente ‘insolent’	4 quedarse - 0 **ponerse - 9**	**93** **volverse - 8** hacerse - 2
31	loco ‘crazy’	**13,319** quedarse – 44 **ponerse - 159**	1,818 **volverse – 3,303** hacerse - 15
32	nervioso ‘nervous’	**3,472** quedarse – 4 **ponerse – 1,672**	432 volverse – 11 hacerse - 0
33	pensativo ‘pensive’	**78** **quedarse - 327** ponerse – 15	17 volverse - 2 hacerse - 0
34	perezoso ‘lazy’	34 quedarse - 1 ponerse – 3	**838** **volverse - 67** hacerse - 13
35	perplejo ‘perplexed’	**377** **quedarse - 293** ponerse - 0	9 volverse - 0 hacerse - 0
36	preñada ‘pregnant’	**597** **quedarse - 68** ponerse - 0	67 (most passive) volverse - 0 hacerse - 0
37	quieto ‘still’	**2,238** **quedarse – 1,485** ponerse - 2	82 volverse - 1 hacerse - 0

38	responsable ‘responsible’	34 quedarse - 3 ponerse - 1	**55,539** **volverse - 32** **hacerse – 22,024**
39	sorprendido ‘surprised’	**2,835** **quedarse – 522** ponerse - 0	5,423 (most passive) volverse – 0 hacerse - 0
40	sumiso ‘submissive’	67 quedarse - 1 ponerse - 2	**1,103** **volverse - 12** hacerse - 4
41	terco ‘stubborn’	7 quedarse - 0 **ponerse - 42**	**826** volverse - 6 hacerse - 0
42	tozudo ‘stubborn’	0 quedarse - 0 ponerse - 2	**382** volverse - 0 hacerse **-** 0

There are also adjectives with a very high number of occurrences with *ser* or *estar*, but with a relatively low occurrence with their corresponding verbs. For example, *esquivo* ‘evasive’ and *sumiso* ‘submissive’ have a much higher number of occurrences with *ser* than with *estar*, but have a relatively low occurrence with *volverse/hacerse*, while *fascinado* ‘fascinated’ has many more occurrences with *estar* and relatively few occurrences with *quedarse/ponerse*. For adjectives with *ser*, this may happen because *ser* implies inherent, long-term qualities that are not likely to change. The same reason could hold within analyses of *ser/estar* in terms of an aspectual distinction, such as Luján (1981) and Clements (2006), where either *imperfective* or [-aspect] correspond to *ser* (no change), and *perfective* or [+aspect] correspond to *estar* (there is a change). Thus, speakers may not feel the need to use verbs of change of state with adjectives that occur with *ser*.

For *estar*, it is not clear why this would happen. However, it is clear from the data in Tables 2, 3 that, overall, verbs of becoming are used more often with adjectives that are used mostly with *estar* (e.g., *asombrado* ‘amazed,’ *nervioso* ‘nervous’) than with *ser*. Of the adjectives that confirm the trend, there is almost double the number of adjectives that occur mostly with *estar* and also with *quedarse/ponerse* than the ones that occur mostly with *ser* and with *volverse/hacerse* (21 and 11, respectively). Since *estar* implies temporary, short-term conditions or states that can change frequently, it stands to reason that verbs of becoming will tend to be used more often with adjectives that are used mostly with *estar*.

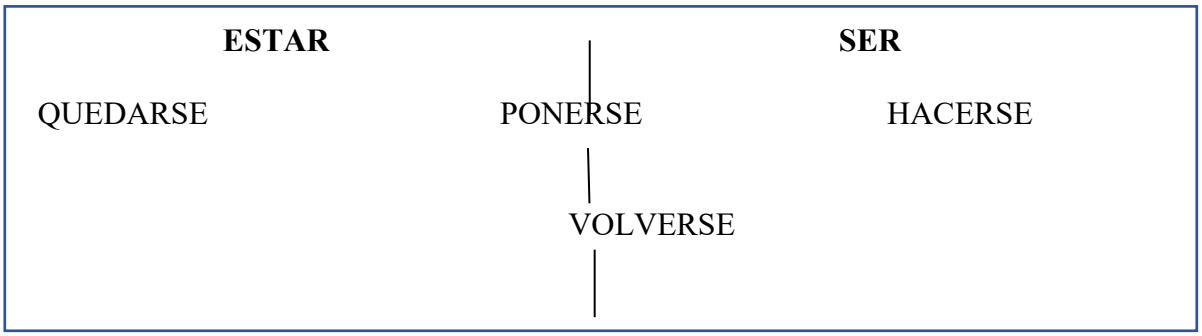
There may be another reason why speakers may not feel the need to use verbs of becoming with certain adjectives. Some adjectives may not be used much with any of the four target verbs because they already have an associated verb that carries the meaning of *becoming*, in a situation reminiscent of blocking. For instance, to say that someone has become *delgado* ‘thin,’ one usually says *adelgazar* ‘to become thin(ner),’ so speakers may not feel the need to use any of the four verbs of becoming to express the change of state. Still, in such cases, the correlation between the use of *ser/estar* and a corresponding verb of becoming persists. Adjectives that have more tokens with *estar* tend to be used more often with *quedarse/ponerse*, while adjectives that have more tokens with *ser* tend to be used more with *volverse/hacerse*. There are other adjectives, such as *desnutrido* ‘malnourished’ and *indignado* ‘indignant,’ which occur frequently with *ser* or *estar*, but not very much or not at all with any of the verbs of becoming. Again, this may be due to speakers seeing no need to express a change of state with these particular adjectives.

Notice in Table 3 that for a few cases, all or most of the uses with *ser* found in the corpus are past participles used in passive sentences (e.g., *fueron asustados* 'they were frightened (by someone or something)'). Because past participles used in passive sentences are verbs and not adjectives (thus they cannot be counted as part of the data), this means that for these instances, their use with *ser* as prototypical adjectives (rather than as participles) was close to nil. By rarely appearing with both the verb *ser* and with *volverse/hacerse*, these adjectives offer further confirmation of the general trend.

As noted in § 2, B&E could not form clusters with any of the adjectives used with *volverse* and *hacerse*, and they could not either with some adjectives with *quedarse* and *ponerse*. All these were designated by B&E as isolated cases. In contrast, the cluster conditions can account for most of the isolated instances that are unrelated to an exemplar in B&E's model, including most adjectives used with *volverse* and *hacerse*, and the adjectives with *quedarse* and *ponerse* that were isolated in B&E.

The diagram in (3) captures in visual form the interaction between *ser* and *estar* and verbs of becoming, while at the same time reflecting the positioning of the verbs in the cluster conditions scale (Table 1). For example, note from (3) how *quedarse* and *hacerse* are on opposite ends, as in the scale. Moreover, as noted above, the scale accounts for overlap in the selection of adjectives, and shows that the overlap is limited and systematic: there is overlap between adjectives that fall nearby in the scale, but there is hardly any overlap when adjectives are on opposite ends of the scale. This happens because adjectives that are closer to each other on the scale are also closer in the number of conditions they satisfy. In (3), this overlap is seen more clearly with *ponerse* and *volverse*, but the diagram also shows that *quedarse* and *ponerse*, on the one hand, and *volverse* and *hacerse*, on the other, each occupy their own "copular" space, with *estar* and *ser*, respectively, and thus each pair of verbs can more easily undergo overlap among themselves in the use of adjectives.

(3) Interaction between *ser* and *estar* and verbs of becoming



A related issue that may be elucidated by the diagram in (3), in conjunction with the scale of cluster conditions, is the question of why there are adjectives that are supposed to be used with both *quedarse* and *ponerse*, on the one hand, or with both *volverse* and *hacerse*, on the other, but only occur with one or the other verb in the pair. For example, as seen from Table 3, *perplejo* 'perplexed,' which is used mostly with *estar*, is used 293 times with *quedarse*, but zero times with *ponerse* (e.g. *se quedó perplejo*/**se puso perplejo*), and *perezoso* 'lazy,' which has 67 occurrences with *volverse*, is not used at all with *hacerse*.

The separation in copular space between *quedarse* and *ponerse* and between *volverse* and *hacerse* in (3) above gives us a visual hint, but the cluster conditions scale (Table 1) can provide

the answer with more precision. It may be that, in the minds of speakers, *perplejo* does not satisfy any of the conditions, not even to a minimal degree, so speakers (unconsciously) decide to use this adjective only with *quedarse*, which is at the lowest end of the scale. Conversely, speakers may feel that *perezoso* satisfies the conditions to a large, but not to the full, extent, so that they sense that using *volverse*, but not *hacerse* (which is at the highest end), is appropriate. Another explanation may have to do with conventionalization, which is exemplified below with the exceptional cases.

The corpus results provide further support for the diagram in (3). For example, the adjectives *agresivo* 'aggressive' and *insolente* 'insolent' show overlap between *ponerse* and *volverse*. As seen in Table 3, each of these adjectives has about the same number of uses with *ponerse* and *volverse*, but in the opposite ends of the scale (*quedarse* and *hacerse*) there are hardly any instances. Thus, with the help of the cluster conditions scale, the overlap in the use of adjectives can be predicted with a relatively high degree of certainty, and this is reflected in the diagram in (3).

Finally, adjectives that, unexpectedly, clearly pattern with their non-corresponding verbs of becoming could be considered exceptions. That is, these adjectives are used predominantly with *ser* or *estar*, but occur mostly with *quedarse/ponerse*, on the one hand, or *volverse/hacerse* on the other hand. As seen in Tables 2, 3, these adjectives are *calvo*, *loco*, and *terco*. Of these, only *calvo* is an isolated case in B&E; *loco* appears in a different cluster with *quedarse*, and *terco* was not an adjective examined in B&E. The explanation for why these are exceptions in the current study could have to do with the fact that, as B&E note, some adjectives are used in a formulaic or highly conventionalized way.

For example, *calvo* and *loco*, which occur mostly with *ser* and *estar*, respectively, occur more frequently with a non-corresponding verb: *calvo* occurs mostly with *quedarse*, and *loco* with *volverse*. This could well be because *quedarse calvo* 'become bald' and *volverse loco* 'go crazy' at some point in their history became conventionalized uses, and that has not changed through time. Thus, the four adjectives found in the present study that buck the trend could be considered exceptions not because they are isolated cases, but rather because of the strength of conventionalization. However, in spite of the fact that, as noted by B&E, conventionalized uses are frequent, the trend based on *ser/estar* demonstrated in this study still holds. This means that it is a robust trend, and conventionalization only affects it in a minimal way.

4.2. Comparison with the Exemplar Model

In several parts throughout the paper, especially in § 2, brief comparisons have been made between the current model, based on cluster conditions, and B&E's exemplar model. This section brings all those observations together and adds relevant details, in order to focus on the advantages of the current model with regards to the exemplar model of B&E, as well as to explore points of convergence between the two models. Before delving into the comparison, we provide an outline of the basic foundations of usage-based linguistics (of which the exemplar model is a part), as well as cognitively-based constructionist approaches, which are likewise usage-based.

A relatively recent overview of usage-based theory is provided by Bybee (2013). She notes that the basic premise of usage-based theory is that experience with language creates the cognitive representations for language. These representations are built up by language users on the basis of phonetic form, meaning and context. Similarity is key, as incoming utterances are matched and organized by their resemblance to existing representations. As this happens, units such as syllables, words, and constructions emerge. For their part, Masini & Audring (2019) establish a connection

between constructions—stored pairings of form and meaning—and usage-based theory, as well as with exemplar-based models. As with these theories, Masini & Audring (2019) assume that grammar is built up and acquired bottom-up from complex words and phrases found in the input. The resulting constructions are stored in lexical memory as schemas.

Other examples of usage-based constructions include the reusable phrasal fragments proposed by Thompson (2002). These fragments, such as *I think* in *I think it'll be real interesting*, express the speaker's stance toward the content of a clause. Rather than being seen as predicates that take complements that are subordinate clauses, they are stored and retrieved as formulaic stance markers; hence their status as usage-based constructions. For another example, Boas (2003) proposes resultative mini-constructions within the frameworks of Construction Grammar and Frame Semantics. These are structures in which each particular sense of a resultative verb (e.g. *paint* in *John painted the house green*) constitutes a conventionalized mini-construction, which provides crucial information for the licensing of arguments.

As noted in the Introduction, both B&E's model and the cluster concepts model are dealing with the same type of construction ([verb of becoming + adjective]). However, while in B&E's model the use of the constructions is determined by similarity to an exemplar, in the current model their use is determined by the application of the cluster conditions (preference rules) in rule-like fashion. As seen in § 3, cluster concepts are rule-like in that they involve the satisfaction of conditions in order to reach an interpretation, and this fact is important because it establishes a clear contrast with exemplar-based theories. Furthermore, while usage and pragmatic phenomena, such as conventionalization, may be relevant for the application (or non-application) of the cluster conditions, the cluster concepts model is framed within a mentalist view, where cluster conditions are a part of the speaker's mental representations, and a calculation of the conditions, as in Table 1, may be going on in the speaker's mind. Thus, similarity plays a key role only in the exemplar model.

The issue of abstract features is another important factor with respect to the similarity vs. rules difference between the two models. Recall that B&E argue that novel instances of [verb + adjective] sequences are based on analogies to previous experience, not on rules that refer to abstract features. They also note that the meaning in the exemplar clusters is created by the user's experience and may not be describable by abstract features that provide strict boundaries. However, B&E admit that there are some abstract features that seem to link some uses of the verbs. For example, a sense of "lacking" can be found in the clusters centered around *quedarse solo* 'to end up alone' and *quedarse quieto*, *callado*, *tranquilo*, and so on, 'to become still, quiet, tranquil'. All of these constructions indicate the lack of some property such as movement, company, or agitation. Thus, B&E seem to acknowledge that reference to some abstract notions may unite these and other sets of constructions. However, abstract features are relevant for the exemplar model only in this limited way. This is in contrast to the current model, where abstract features such as "change" and "action" are key for the postulation and application of the cluster conditions.

While the similarity vs. rules contrast is an important difference between the two models, token frequency appears to be a key factor for both of them. While types are important for both models because they determine the number of adjectives to be examined, the number of tokens can be said to be equally, or maybe more important for both models. For the B&E model, tokens are important because adjectives with more tokens are key to form clusters, in that they become the central member of the cluster, while for the current model, the number of adjective tokens determines the correlation between adjectives and verbs of becoming; as noted above, adjectives that have more

tokens with *estar* tend to be used more often with *quedarse/ponerse*, while adjectives that have more tokens with *ser* tend to be used more with *volverse/hacerse*.

Moving on to the comparison between the two models based specifically on the date collected, first, as noted above, even though B&E argue that the use of the copular verbs *ser* or *estar* with a given adjective cannot be used as a reliable criterion to determine what verb of becoming that adjective occurs with, the corpus study results suggest that, in fact, there is a strong correlation between the use of an adjective with *ser* or *estar* and the verbs of becoming the adjective occurs with. As seen in Tables 2 and 3, adjectives that are used mostly with *estar* (e.g., *está perplejo* 'is perplexed') tend to occur more frequently with *quedarse*/*ponerse*, while adjectives that are used mostly with *ser* (e.g., *es arrogante* 'arrogant') tend to occur with *volverse/hacerse* more commonly.

Another important difference is that the cluster concept model accounts for a significant number of uses of adjectives that do not fall into any particular grouping in B&E's exemplar model. Thus, unlike B&E, the cluster concept model can account for the vast majority of uses of adjectives in the target constructions, as seen in Tables 2, 3. Recall that B&E found no exemplar-based clusters for adjectives that occur with *volverse* and *hacerse*, and they also found isolated uses for some adjectives that occur with *quedarse* and *ponerse*. As observed by B&E, this occurred due to lack of similarity to other adjectives. The verbs *volverse* and *hacerse* are used with more diverse sets of adjectives, and B&E found that these adjectives do not fall into clusters. That is, the adjectives that combine with *volverse* and *hacerse* (e.g., *volverse esquivo* 'to become evasive,' *hacerse responsable* 'to become responsible'), as well as some adjectives with *quedarse* and *ponerse* (e.g., *embarazada* 'pregnant,' *pesado* 'annoying'), are semantically isolated uses that do not bear a semantic resemblance to any other exemplars of a given construction. Thus, they are not accounted for in the exemplar model.

In contrast, the current model accounts for the behavior of adjectives, including the "isolated" uses, in all four constructions, on the basis of their position on a scale (Table 1) based on cluster concepts, and their association in meaning with *ser* and *estar*. The cluster conditions can account for most of the isolated instances that are unrelated to an exemplar in B&E's model, including most adjectives used with *volverse* and *hacerse*, and the adjectives with *quedarse* and *ponerse* that were isolated in B&E. For example, the following adjectives were found to be isolated instances by B&E, but fit into the classification provided by the cluster conditions scale: *esquivo* 'evasive,' *responsible* 'responsible,' *embarazada* 'pregnant.' The behavior of adjectives such as *embarazada*, an isolated use in B&E's study, and of *nervioso* 'nervous' and *sorprendido* 'surprised,' for which B&E presented two clusters, turns out to be completely predictable on the basis of their use with *ser*/*estar* and cluster concepts.

Thus, while isolated uses are odd and unaccounted for in the exemplar model, they are naturally accounted for by cluster concepts because they satisfy a certain number of conditions and fall on a specific point in the scale, along with other verbs. What are isolated uses in B&E's study are produced naturally by native speakers and need to be accounted for by any model that aims to explain the behavior of verbs of becoming in a comprehensive way. The current model does this, something that cannot be accomplished by B&E's exemplar-based model.

Regarding adjectives specifically with *quedarse* and *ponerse* that turned out to be isolated uses in B&E's study, consider two examples. First, the conventionalized way of saying 'to become pregnant' in Spanish is *quedarse embarazada*, which occurred four times in the B&E corpus study. Of the other tokens with *quedarse* that also described physical states (e.g., *desnutrido* 'malnourished,' *calvo* 'bald'), none were considered to form a cluster with *embarazada* 'pregnant;'

no cluster figure could be provided by B&E for this group because the adjectives are distributed in separate, nonoverlapping areas of conceptual space. In contrast, in the current study, *embarazada* occurs overwhelmingly with both *estar* and *quedarse*, fitting the trend whereby adjectives that occur mostly with *estar* also occur mostly with either *quedarse* or *ponerse* (or both).

Note that while the number of isolated uses in B&E's corpus study was small, what is crucial is that no clusters at all could be formed with the verbs *volverse* and *hacerse*. Considering that a large number of adjectives occur with these two verbs, this represents a significant gap in the explanatory range of the exemplar-based model. In contrast, there are no gaps in the explanation offered by the cluster concepts model, as it accounts for the vast majority of adjectives used with all four verbs of becoming under study. Including adjectives that confirm the trend (79%) and adjectives in the middle ground (14%) (which, recall, are not exceptions), a total of 93% of adjectives is accounted for by the cluster conditions model.

Notice that, as noted above, the adjectives in the two clusters in Figures 1, 2 from B&E have in common an absence of movement, agitation or company. However, there is a cluster in B&E's study which does not share features with any of the other clusters with *quedarse* (nor is it the opposite of any other cluster). This is the cluster shown in Figure 3, which centers on *sorprendido* 'surprised'. Thus, say B&E, it appears that the various categories of adjectives used with *quedarse* need not be related to one another. However, this means that in the exemplar-based model there are not only isolated instances of adjectives, but also isolated clusters of adjectives. In contrast, as seen above, in the model proposed here, which uses cluster concepts, most adjectives used with *quedarse* are shown to have in common a consistent relation *estar*.

Finally, just like cluster conditions provide flexibility in the determination of word meaning (see § 3), they enable flexibility in the determination of the use of verbs of becoming. As observed in § 2, the overlap in the use of verbs of becoming (e.g., using an adjective with both *quedarse* and *ponerse*) is problematic for models based on categorical criteria because, by definition, categories have boundaries, so any overlap indicates that the category boundaries are not well defined. By contrast, the scale in Table 1, being a continuum with no boundaries, explains the overlap in a systematic way.

In sum, there is a strong correlation between the use of an adjective with *ser* or *estar* and the verb of becoming the adjective occurs with. In addition, both isolated instances and isolated clusters in the B&E model find an explanation under the cluster concept model, which allows for flexibility and avoids the pitfalls of categorical criteria.

4.2.1. Drawbacks of Connectionist Models and Exemplar Models Based on Analogy

In this section and the next I present ideas from several authors that both argue against exemplar models, which are based on analogy, and lend support to the model developed in this study, by favoring rule-like operations to explain linguistic phenomena. As discussed above, the fact that cluster conditions are rules (they are also known as preference *rules*) is relevant because it establishes a clear contrast with exemplar theories, which are based on similarity. This contrast can be seen more clearly when we consider isolated uses. As noted above, in B&E's analysis there are semantically isolated uses that do not bear a semantic resemblance to any other exemplars of a given construction. Thus, they are not accounted for in the exemplar model.

This receives a possible explanation through Pinker's (2006) observation that connectionist models and models based on analogy, both of which, like the exemplar model, are based on similarity, cannot adequately account for cases where no resemblance exists between items. For example, while connectionist/analogical models do a good job of generalizing irregular patterns,

such as converting (the nonce verb) *spling* to the past form *splung* from the similar *cling-clung*, and generalizing some regular-sounding patterns, such as converting *plip* to *plipped* based on similarity to *flipped* and *clipped*, only people, but not the analogical models, could convert *ploamph* (not similar to any existing verb) to *ploamphed* (in a study cited by Pinker). The connectionist model instead produced strange forms such as *ploamph-bro, membled* from *mail*, pairs such as *smeej-leaflog* and *frilg-freezld*, and failed to generate any past forms at all for *jump* and *pump*. In contrast, people reason that a verb is a verb, and they attach the suffix *-ed*, no matter how strange-sounding the verb is. The attachment of *-ed* is the regular rule that applies as a default to items that do not resemble other items (see also Pinker 1999, Pinker & Ullman 2002, Huang & Pinker 2010).

Because irregulars form clusters with similar items that overlap (*sing-sang-sung*; *ring-rang-rung*), and usually do not generalize outside their clusters, analogical and connectionist models, with nothing to fall back on if an item does not overlap in some way with other items, can only produce the bits and pieces that are closest to the ones they have been trained on.

Eddington (2000) proposed an exemplar model based on analogy that arguably avoids the drawbacks of analogy-based models as outlined by Pinker. However, in Eddington's (2000) model there is a large amount of innate machinery, including a similarity algorithm and *subcontexts* and *supracontexts*. The algorithm contains an explicit procedure for assembling an analogical set from which analogs may be chosen, and its purpose is to determine which members of the database are most likely to affect the inflection of a word, say the verb *sin***,** and also to calculate the extent of analogical influence exerted. The *subcontexts* and *supracontexts* make it possible to find disagreements between forms (e.g., between *sin*, *thin* and *sing*). The analogical set contains all of the database items that can possibly influence the regularity of *sin*, an indication that the model is geared to find the regular forms of an item.

This is all very complex innate machinery, which is inconsistent with the central tenets of exemplar models, keeping in mind that a key element of usage-based analogical models, in contrast to rule-based models, is an avoidance of innate constructs. For example, Bybee (2013) observes that in the exemplar-based approach, linguistic structure emerges when human cognition deals with tokens of experience with the world. According to Bybee (2013), in usage-based theories, grammatical processes are part of domain-general cognitive processes which do not depend on innate, abstract, and linguistic-specific grammatical categories. It seems as if the algorithm in Eddington's (2000) model is elaborate enough such that it has been set up to capture regular forms, that is, the rule that produces regulars. This is reminiscent of what Pinker (2006) observes regarding connectionist models, namely, that modelers have hardwired various parts into their model that are tailor-made for regular verbs (p. 225, 233), contravening the idea that the analogical network by itself, with the training it has received, should compute the correct regular and irregular forms.

According to Pinker (2006), the problem of computing coherent past tense forms for novel-sounding verbs still has no satisfactory solution in connectionist models and, by extension, in analogical models (such as Eddington's 2000). In contrast, the dual route model makes use of very little machinery; a simple concatenative rule (e.g. add *-ed* for the past tense) for regulars and association by similarity for irregular forms.

An interesting middle ground between analogy and rules is presented by Albright & Hays (2003), who argue against analogy but propose a model that uses multiple induction-based rules based on phonological contexts rather than a default rule for regular forms. These rules are assigned reliability scores according to their performance in the existing lexicon. Albright & Hays

(2003) argue that learners of English go beyond a single default rule for regulars, and learn a set of detailed environments that differentiate the degrees of confidence for the regular outcome. Albright & Hays (2003) do share with the mainstream tradition of formal linguistic theory the view that linguistic knowledge is best characterized by rules. They note that because rules contain variables, they permit correct outputs to be derived even for unusual input forms that lack neighbors.

Further discussion of these studies is beyond the scope of the present paper, but future research could address these issues in order to establish a clearer comparison between models based on analogy and those based on rules or rule-like mechanisms, such as the cluster conditions model. What is important from the discussion is that in arguing against purely analogical models, it lends support to the model based on cluster conditions developed in this study, which is a plausible alternative to models based on analogy.

4.2.2. Additional Drawbacks of Exemplar Theories

While Jackendoff (2002, 2019) and Jackendoff & Audring (2020) embrace the notion of *construction* as discussed above (see also Goldberg & Jackendoff 2004), Jackendoff & Audring (2020) put forward further shortcomings of exemplar models, and, by extension, of usage-based theories. Jackendoff & Audring (2020) present a stance of strong doubt and incredulity when it comes to the claims and assumptions of exemplar theories. They point out that a variety of exemplar theories posit that one's brain stores every input one has experienced, and some approaches (e.g. Bybee 2001) claim that token inputs are all that one stores. A first response to these theories from Jackendoff & Audring (2020) is incredulity, because this implies that one stores a memory of every cup of coffee one has ever drunk, of every doorknob one has ever turned, of every time one has shaken hands with someone, of every time one has said *good morning* or *the*, and so on. For Jackendoff & Audring this seems wildly excessive and implausible.

A second concern for Jackendoff & Audring (2020) involves what may be called *granularity*. What size units does one store? For example, if one stores entire sentences and discourses, does one also separately store every single token of their words, every single token of their phonological segments, and/or every acoustic trace of the utterance, including speaker's tone of voice, all in full detail? This seems, again, highly implausible. With regard to lexical storage, Jackendoff & Audring (2020) cite McQueen, Cutler, and Norris (2006) on experimental evidence against exemplar theories of word learning and in favor of abstractionist accounts, where lexical representations help categorize lexical knowledge.

Moreover, according to Jackendoff & Audring (2020), for those versions of exemplar theory that deny the existence of abstract categories (such as B&E's and Bybee's 2013), there is a further logical problem: how does one store experiences? You cannot simply copy sunsets, horses or vowels into your head. There has to be a format, template or schema for storage—some repertoire of features and/or analog dimensions of similarity and variation, in terms of which one's experiences are encoded. However, a feature defines a category of things that have that feature, and a dimension of variation likewise is an abstraction over the tokens. In other words, according to Jackendoff & Audring (2020), a category- and abstraction-free coding of experience is impossible.

Finally, Jackendoff & Audring (2020) also find exemplar theory implausible as a model of processing, both in general, and more specifically with respect to the speed of lexical access during processing. Exemplar theories such as B&E's and Bybee's (2013) claim that the brain stores every single instance of a word's use, building up a mental corpus for which the written corpus is a

proxy. The idea, then, is that more frequent words produce more mental tokens, which then combine forces to strengthen or speed up response. However, Jackendoff & Audring point out that it is not clear how to apply this approach to language production: how does a speaker construct an output based on a cloud of exemplars (such as B&E's clusters of related adjectives illustrated above)? And, in either production or comprehension, are all the exemplars activated? It would seem to take more work rather than less to activate thousands or tens of thousands of stored instances (as opposed to categories of items), limiting the speed of lexical access during processing. B&E cite analogy as the mechanism that accounts for productive use, but, as noted above, they do not show explicitly how clustering can predict subsequent uses of a given construction, thus not providing an answer to Jackendoff & Audring's question of how a speaker can construct an output based on a cloud of exemplars.

While, again, it is beyond the scope of the current article to analyze studies that provide arguments that counter Jackendoff & Audring's (2020) claims (such as Pierrehumbert (2001) and Ambridge (2020)), by highlighting the disadvantages of exemplar models, the brief discussion above lends further support to the model based on cluster conditions developed in this study, which provides an alternative mechanism to account for the behavior of verbs of becoming.

5. Future Lines of Research: Application to Other Phenomena and to Processing

The cluster concept scale (Table 1) could be applied to other phenomena where all-or-none rules do not work, such as the "personal *a*" in Spanish, an instance of differential object marking (see Bossong 1991), where several factors such as animacy, specificity, definiteness, and possibly agency seem to play a role in determining when the preposition *a* must precede an object argument. For example, the "personal *a*" occurs in *Vi a Pedro* 'I saw Pedro' (animate object) and is absent in *Vi la silla* 'I saw the chair' (inanimate object). However, animacy does not seem to be the only factor in its application. While one can say *Tengo a la hermana de Pedro en mi lista* 'I have Pedro's sister in my list,' and *Conozco al Doctor Pérez* 'I know Doctor Pérez,' one cannot say **Tengo a una hermana* 'I have a sister' or **Conozco a un buen doctor* 'I know a good doctor' (although this sentence may be acceptable in some dialects). The latter two sentences are usually grammatical only without the "personal *a*." If we take animacy and specificity/definiteness as two different conditions, we can account for the ungrammaticality of sentences such as **Tengo a una hermana* by saying that while the animacy condition is satisfied, the specificity/definiteness condition is not, in contrast to *Vi a Pedro*, where both conditions are satisfied. Note that verbs such as *asustar* 'frighten' and *ayudar* 'help' are obligatorily followed by *a* and only take animate objects, unlike verbs such as *ver* 'see,' which take both animate and inanimate objects. Thus, the cluster conditions for the "personal *a*" would not be relevant to the former kind of verbs.

We can also envision a realization of the cluster conditions scale in terms of processing. After all, cluster conditions presumably are used by native speakers (unconsciously) when determining the meaning of words (Jackendoff 2002). In this mentalist view, as noted above, cluster conditions are a part of the speaker's mental representations, and a calculation of the conditions, as in Table 1, may be going on in the speaker's mind. This would be parallel to a situation where, for instance, Spanish native speakers have to determine, based on several internalized semantic (and syntactic) criteria, whether to use the preterite or imperfect, *ser* or *estar*, or the subjunctive in a given context. For example, say a native speaker of Spanish (who does not speak English) wishes to express an idea such as "Peter was nervous all night at the party." The speaker needs to decide whether to use *ser* or *estar*, and whether to use the preterite or the imperfect. It may well be that speakers are making a calculation (based on the conditions for *ser*/*estar* and preterit/imperfect) using the

mechanism of cluster conditions. In this example, it turns out that *estar* needs to be used, not *ser*, and the verb needs to be in the preterite. The result is the sentence *Pedro estuvo nervioso toda la noche en la fiesta*.

The use of cluster conditions may even help explain marginal uses of *ser* or *estar*, such as cases where, say, *ser* is used many more times than *estar* with a given adjective, but *estar* should presumably not be used at all (e.g., *están sumisos* 'they are submissive' is a rare usage; see Tables 2, 3. It could well be that speakers focus on one or two conditions for one of these copular verbs, determine (unconsciously) that these conditions fit the situation, select that verb (say, *estar*), and do not bother to compute for the remaining conditions. The verb selected could thus appear in a non-prototypical usage.

The process may take place in a similar way when speakers make computations to select a verb of becoming. Using the cluster conditions scale, speakers have an adjective in mind and need to select a verb of becoming that can best combine with that adjective. Some speakers then tick off conditions and select a verb for the adjective they have in mind once a certain number of conditions has been met, while other speakers may go through all the conditions and select the verb that satisfies most of them.

6. Conclusion

We have seen how the notion of cluster concepts instantiated in a cluster conditions scale accounts for the behavior of Spanish verbs of becoming in a more precise and comprehensive way than a usage-based approach. The tendencies in the use of adjectives with a given verb of becoming, which show a strong correlation to their use with *ser/estar*, are formalized through the use of the cluster conditions scale, which allows for flexibility and is not based on categorical, all-or-none judgments. In sum, what emerges is a scale of cluster concepts that motivates the correlation between *ser/estar* and verbs of becoming in Spanish.

Part of our mental lexicon is organized in an associative network (Pinker 1999, 2006) based on similarity, and the exemplar-based approach does capture clear-cut cases of clustering that are consistent with this lexical organization. However, there are uses of adjectives in constructions with verbs of becoming that do not fit any specific pattern, and cannot be captured by the usage-based approach. Specifically, this is due to a lack of semantic resemblance to other adjectives, which forces the exemplar-based model to classify these adjectives as isolated uses. In contrast, the model proposed here, which is based on cluster conditions set on a scale, coupled with an interaction with the verbs *ser/estar*, can account for and predict the use of adjectives that occur with all four verbs of becoming. Crucially, while the current model accounts for all four constructions under investigation (with *quedarse*, *ponerse*, *volverse* and *hacerse*), B&E's exemplar model only accounted for two (*quedarse*, *ponerse*).

Future studies could examine how well the cluster concept model can account for other change of state verbs such as *convertirse*, *transformarse* and *llegar a ser*, which occur mostly with nouns. Further research could also explore the use of verbs of becoming with a wider number and range of adjectives, preferably using a large corpus, as in this study. Also to be explored in more detail is whether other grammatical phenomena such as the "personal *a*," *ser*/*estar*, and preterite/imperfect are amenable to explanation through the use of the cluster concept scale during processing, as well as how the notion of cluster concepts can be incorporated into language instruction.

Finally, in English, some of the equivalents for Spanish verbs of becoming are *to become*, *to turn*, *to end up*, *to get*, and *to go*. As in Spanish, certain adjectives in English cannot occur with

certain verbs of becoming (cf. *became ill* vs. **turned ill*; *got hurt* vs. **became hurt*; *turned red* vs. **got red*; *went crazy* vs. **went nervous*). As part of comparative research, it would be interesting to investigate if and how the notion of cluster concepts applies to these English verbs; the results could help us further understand both the cluster conditions scale and Spanish verbs of becoming.

Carlos Benavides
University of Massachusetts Dartmouth
285 Old Westport Road, Dartmouth MA 02747
cbenavides@umassd.edu

References

Albright, A., & Hayes, B. 2003. Rules vs. analogy in English past tenses: A computational/ experimental study. *Cognition* 90(2).119-161. https://doi.org/10.1016/S0010-0277(03)00146-X

Ambridge, Ben. 2020. Against stored abstractions: A radical exemplar model of language acquisition. *First Language* 40(5-6).509-559. https://doi.org/10.1177/0142723719869731

Boas, Hans. 2003. *A constructional approach to resultatives*. Stanford: CSLI Publications.

Bossong, Georg. 1991. Differential object marking in Romance and beyond. In D. Kibbee and D. Wanner (eds.), *New analyses in Romance linguistics*, 143–70. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.69.14bos

Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511612886

Bybee, Joan. 2013. Usage-based theory and exemplar representations. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford handbook of Construction Grammar*, 49-69. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195396683.013.0004

Bybee, Joan, and David Eddington. 2006. A usage-based approach to Spanish verbs of becoming. *Language* 82.323-355. https://doi.org/10.1353/lan.2006.0081

Clements, J. Clancy. 2006. *Ser-estar* in the Predicate Adjective Construction. In J. Clancy Clements & Jiyoung Yoon (eds), *Functional Approaches to Spanish Syntax*, 161-202. Basingstoke UK, Palgrave-Macmillan, London. https://doi.org/10.1057/9780230522688_8

Davies, Mark. 2002. *Corpus del Español*. www.corpusdelespanol.org.

Delbecque, Nicole and Lise Van Gorp. 2015. The pseudo-copular use of the Spanish verbs *hacerse* and *volverse*: two types of change. *CogniTextes* Vol. 13. DOI: https://doi.org/10.4000/cognitextes.843.

Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67.547-619. https://doi.org/10.1353/lan.1991.0021

Eddington, David. 1999. On 'becoming' in Spanish: A corpus analysis of verbs expressing change of state. *Southwest Journal of Linguistics* 18.23-35.

Eddington, David. 2000. Analogy and the dual-route model of morphology. *Lingua*, 110(4), 281-298. https://doi.org/10.1016/S0024-3841(99)00043-1

Eddington, David. 2002. Disambiguating Spanish change of state verbs. *Hispania* 85.921-29. https://doi.org/10.2307/4141261

Goldberg, Adele, & Ray Jackendoff. 2004. The English resultative as a family of constructions. *Language* 80: 532-568. https://doi.org/10.1353/lan.2004.0129

Huang, Yi Ting, and Steven Pinker. 2010. Lexical semantics and irregular inflection. *Language and cognitive processes* 1-51. https://doi.org/10.1080/01690961003589476

Jackendoff, Ray. 1983. Semantics and cognition. Cambridge, MA: MIT Press.

Jackendoff, Ray. 2002. Foundations of language: Brain, meaning, grammar, evolution. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198270126.001.0001

Jackendoff, Ray. 2019. Conceptual Semantics. In Claudia Maienborn, Klaus von Heusinger, and Paul Portner (eds.), *Semantics: Theories*, 86-113. Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110589245-004

Jackendoff, Ray, and Jenny Audring. 2020. *The texture of the lexicon: Relational Morphology and the Parallel Architecture*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198827900.001.0001

Luján, Marta. 1981. The Spanish copulas as aspectual indicators. *Lingua* 54.165-210. https://doi.org/10.1016/0024-3841(81)90068-1

Marcos Marin, Francisco. 1992. Corpus oral de referencia del español contemporáneo. Textual corpus, Universidad Autónoma de Madrid.

Masini, Francesca, & Jenny Audring. 2019. Construction Morphology. In Jenny Audring & Francesca Masini (eds.), *The Oxford handbook of morphological theory*, 365-89. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199668984.013.25

McQueen, Cutler, and Norris. 2006. Phonological Abstraction in the Mental Lexicon. *Cognitive Science* 30.1113-1126 . https://doi.org/10.1207/s15516709cog0000_79

Morimoto, Yuko, and Victoria Pavón Lucero. 2004. Aproximación semantica a la gramática de *ponerse* y *quedarse*. *Studia Romanica Posnanensia* 31.385-392. https://doi.org/10.14746/strop.2004.31.037

Pierrehumbert, Janet. 2001. Exemplar dynamics: Word frequency, lenition and contrast. *Frequency and the Emergence of Linguistic Structure* 45.137-157. https://doi.org/10.1075/tsl.45.08pie

Pinker, Steven. 1999. Words and Rules. New York: Basic Books.

Pinker, Steven. 2006. Whatever happened to the past tense debate? *Wondering at the Natural Fecundity of Things: Essays in Honor of Alan Prince*, ed. by Eric Bakovic, Junko Ito, John J. McCarthy. Createspace Independent Pub.

Pinker, Steven, y Michael T. Ullman. 2002. The past and future of the past tense. *Trends in Cognitive Sciences* 6.456-463. https://doi.org/10.1016/S1364-6613(02)01990-3

ten Hacken, Pius. 2014. Delineating derivation and inflection. The Oxford Handbook of Derivational Morphology, ed. by Rochelle Lieber and Pavol Štekauer. Oxford: Oxford University Press, 10-25.

Thompson, Sandra. 2002. Object complements and conversation: Towards a realistic account. *Studies in Language* 26. 125-164. https://doi.org/10.1075/sl.26.1.05tho

Van Gorp, Lise. 2015. Pseudo-copular use of the Spanish verbs *ponerse* and *quedarse*: two types of change. *CogniTextes* Vol. 13. DOI : https://doi.org/10.4000/cognitextes