# Frequency effects and the study of Limon's Spanish copulas: ser and estar

**Por: Jorge Aguilar-Sánchez[1], Universidad Nacional, Costa Rica, ORCID: 0000-0003-4457-2767**

---

1    Jorge Aguilar-Sánchez es bachiller en la Enseñanza del Inglés de la Universidad Nacional; también es doctor en Lingüística y Lingüística Hispánica, máster en Lingüística Hispánica y en TESOL y Lingüística aplicada por la Universidad de Indiana. Trabaja como docente en la Escuela de Literatura y Ciencias del Lenguaje de la Universidad Nacional, Costa Rica. Contacto: jorge.aguilar.sanchez@una.cr

3    Merriam-Webster. (n.d.). Thesaurus. In *Merriam-Webster*. https://www.merriam-webster.com/thesaurus/thesaurus

## Resumen

**Los efectos de la frecuencia y el estudio de las cópulas en español en Limón: ser y estar**

Este trabajo presenta un resumen comprensivo de lo que se ha hecho con respecto a la cópula del español. Explora cómo los efectos de frecuencia, basados en el Modelo ejemplar, pueden ayudar a descubrir la variación no explicada que se encuentra en los diferentes modelos propuestos. Se busca proponer un marco metodológico para el estudio de la cópula del español en situaciones de contacto, el cual incluye la frecuencia como una variable para entender si el cambio es continuo o estable, y si es de *estar* extendiéndose a los contextos de *ser*, de *ser* extendiéndose a los contextos de *estar* o ambos. Exploro cómo cada uno de los componentes del esquema interactúan entre sí; así mismo explico cómo estas interacciones encajan dentro de la descripción tradicional de la selección de cópula, y me alejo de la visión generativa del lenguaje como un grupo de reglas, para empezar a verlo como un sistema altamente afectado por la frecuencia del uso de sus componentes. Por último, propongo un procedimiento sobre cómo explicar el comportamiento de la cópula en español, que incluye un análisis multinivel de este fenómeno.

## Abstract

This paper presents a brief but comprehensive review of what has been done regarding the Spanish copula. It explores how frequency effects can help account for the unexplained variance found in the different proposed models, based on the Exemplar Model. This is done to propose a theoretical framework for the study of the Spanish copula in contact situations that include frequency as a variable to understand whether change is ongoing or stable, and whether the change of *estar* is extending to the contexts of *ser*, *ser* is extending to the contexts of *estar*, or both phenomena is taking place. I explore how each of the schema components interacts with each other, how these interactions fit into the traditional description of copula choice, and move away from the generative view of language as a set of rules to begin looking at it as a system highly affected by the frequency of use of its components. I then propose a procedure on how to explain copula behavior in Spanish that includes a multi-level analysis of this phenomenon.

### INTRODUCTION

The study of copula choice in Spanish in contact situations (i.e. among Spanish-English bilinguals) in contexts where Spanish is a minority language (i.e. Southwestern United States) has suggested that copula use is undergoing change. Part of this change involves that the copula *estar* is becoming more accepted in an extended pre-adjectival context, especially with adjectives of *size*, *physical appearance*, age, and those of *description* and *evaluation*. This change has been attributed to contact with English (Silva-Corvalán, 1986 & 1994b)., However, this claim has been challenged by Gutiérrez (1992 & 1994) because he found that monolingual Mexicans living in Mexico and belonging to the same social class and geographical region as those studied by Silva-Corvalán (1986), showed the same behavior regarding copula choice. Diaz-Campos & Geeslin (2011) found similar results concerning copula choice in an analysis of the spoken Spanish of Caracas, a language with no direct contact with English. More recently, Geeslin and Guijarro-Fuentes' (2006) work

on the acquisition of Spanish by native speakers of Portuguese and Aguilar-Sánchez (2009, 2012 & 2017) studied this phenomenon in a context where English is a minority language and Spanish a majority language (i.e., Costa Rica). They found similar results as those provided by Díaz-Campos and Geeslin (2011). Aguilar-Sánchez (2007), however, studied this phenomenon with data from the Corpus de Habla Culta[3] (1987) and Aguilar-Sánchez (2009, 2012 & 2017) in the Spanish-English contact found in Limón, Costa Rica, an area of research requiring further exploration.

Furthermore, all the studies mentioned above have not explained how frequency effects can help account for the unexplained variance found in the different proposed models. In this paper, I explore frequency effects based on the Exemplar Model (Bybee, 2001 & 2002). To do so, I consider the frequency of the two copulas and their variables to determine whether frequency helps explain the existing Spanish variation in copula choice. Therefore, the main goal is to provide a theoretical

---

3    Learned Speech Corpus.

framework for the study of the Spanish copula in contact situations that can be later tested through empirical studies.

Because the scope of this paper is limited, I present a brief but comprehensive review of what has been done regarding the Spanish copula and copula choice. This review is divided into two sections: theoretical accounts and empirical studies of copula choice.

## THEORETICAL ACCOUNTS

Early studies on copula distinction date back to the late 19th and early 20th centuries. During that time, *ser* was defined as a predicate of a permanent condition and *estar* as a predicate of a non-permanent condition. Andrade (1919) associates *estar* with characteristic feelings of immediate perceptions and their representations, and *ser* with objective *description*s. Another descriptive work, Morley (1925), proposes that the contrast between *ser* and *estar* is usually *durative* and *transitory*, or between *act* and *state*.

On the other hand, Bolinger (1944) proposes that the nature of the subject is an important feature in the selection of *ser* and *estar*. For instance, abstract nouns require *ser* (e.g. La santidad es un don 'Sanctity is a gift'). However, abstract nouns referring to concrete experiences might be used with *estar* (e.g. La santidad está decadente hoy en díaa 'Sanctity is decadent nowadays').

Furthermore, the notion of comparison has played a role in the description of copula choice in Spanish. It refers to whether there is a comparison between the referent and him/herself at a different period of time or between him/herself and a group (Franco & Steinmetz, 1983 & 1986). The selection of *ser* + *adjective* and *estar* + *adjective* could be based on subjective factors and the perception of the external phenomenological world. Franco and Steinmetz (1983 & 1986) emphasize that *estar* is not the only copula that indicates a comparison or contrast. They claim that either copula can indicate a comparison and/or a contrast. Thus, *ser* makes a comparison between the referent and a group whereas *estar* compares the referent with him/herself at a different point in time.

Another distinction between copulas is put forward by De Mello (1979). He states that when analyzing copula choice in Spanish, we need to consider the three functions of *ser* and *estar*. These verbs function as main verbs (i.e., action verbs), auxiliary verbs, and attributive verbs (i.e., as copulas). De Mello (1979) states that both copulas have semantic value when used as main verbs; however, as attributive verbs, *ser* loses its semantic value and *estar* acquires a meaning that goes beyond its function as an attributor. DeMello (1979, p. 338) categorizes auxiliary verbs *ser* + past participle as *passive voice* and *estar* + past participle as a *resultant state*.

Carlson (1977) and Kratzer et al. (1995) propose the *Stage-level/Individual-level* hypothesis. This hypothesis states that the copula *ser* combines with individual-level predicates (ILPs) to partially express permanent or essential properties such as *inteligente* "intelligent". It also states that *estar* combines with *stage-level* predicates (SLPs) to express temporary or accidental properties such as *cansada* "tired-FEMENINE". Maienborn (2005) argues against the *Stage-level/Individual-level* hypothesis, stating that *ser* and *estar* both display the same lexical and semantic properties and that *estar* differs from *ser* only in presupposing a relation to a specific discourse situation. Thus, *estar* is the discourse-dependent variant of *ser,* This discourse dependency is lexically initiated by *estar,*structurally resolved by its aspect, and pragmatically licensed through some kind of topic situation contrast. Maienborn (2005) claims that by using *estar* a speaker restricts his/her claim to a specific discourse situation, whereas the speaker makes no such restriction by using *ser*. Clements (2006) states that instead of being predicate types as defined by the *Stage-level/Individual-level* hypothesis, there are two readings that the copula use gives to the *copula + adjective* construction. The first one is an *individual-level* reading which refers to a defining characteristic of the referent. The second is a *stage-level* reading which is a comparison of the characteristic of the referent, not against others, but against itself at an earlier time.

Givón (1984 & 2001) affirms that adjectives come in antonymic pairs. These pairs can either be in a gradual-degree relation to one another (e.g., *seco*

"dry" – *mojado* "wet") or in an absolute relation (e.g., *vivo* "alive" – *muerto* "dead"). The adjective type *seco* "dry"- *mojado* "wet" corresponds to activity and stative verbs, according to Vendler's (1957) classification of verbs. The type *vivo* "alive" – *muerto* "dead" corresponds to non-graduating states which can be analogous to the notion of achievements. Thus, Givón (1984, 2001) proposes the term *gradiency* as an adjective category which refers to the classification of adjectives in two groups. One group is that of non-gradient adjectives or absolute adjectives (i.e., adjectives that cannot be found in "more or less" constructions such as *casado* "married") while the other is gradient adjectives (i.e., adjectives that can be found in "more or less" constructions such as *mojado* "wet").

In sum, theoretical accounts of copula choice have yielded the foundations to its study in Spanish. These foundations have helped different researchers in the analysis of data to explain the choice speakers make between *ser* and *estar*. I now turn to a review of studies that have put these theoretical accounts to the test with empirical data from different sources, including Spanish in contact, Spanish second language acquisition, and monolingual Spanish. These studies have yielded different models for the explanation of copula selection in Spanish. Furthermore, they have produced the variables in the study of copula choice that I will use to propose a model that includes frequency as part of the analysis.

## PREVIOUS MODELS FOR THE STUDY OF COPULA CHOICE IN SPANISH

Second Language Acquisition of Spanish and Sociolinguistics are two fields that have experienced a significant growth in the development of models to explain the use of copula choice. In general, these studies have provided two distinct hypotheses. The first states that a change is ongoing (Silva-Corvalán, 1986) while the second that it is a stable situation (Díaz-Campos & Geeslin, 2011). This hypothesis will be developed in the following section. This section is divided into studies of Spanish-English contact, Spanish second language acquisition, and studies of monolingual Spanish.

### Spanish in Contact with English

Silva-Corvalán (1986 & 1994b) explored the extension of *estar* in the speech of 27 bilinguals of different generations and degrees of Spanish language attrition. There were three groups of participants (GI (immigrants), GII (immigrants under 6), and GIII (Mexican-Americans)). Silva-Corvalán (1986 & 1994b) studied all contexts of *ser* and *estar* and claims that the innovation found in her studies represent part of an evolutionary trend in Spanish and other Romance languages where contact accelerates this trend. She states that the condition of reduced access or lack of access to formal varieties of the language promotes its diffusion and internally motivates changes involving generalizations across languages, which are further accelerated in a situation of extended contact.

With regard to copula choice, Silva-Corvalán (1986, p. 604) suggests that the extension of *estar* in progressives (i.e., *estar* + present participle) and its frequent association with *be* (i.e., *be + verb-ing*) in these constructions may favor the rapid diffusion of *estar* in predicate adjectives, where Spanish has evidenced a slow process of change in spite of any language-specific influence. She also suggests that there is no noticeable movement toward a steady functional specialization, but that the continuous renovation of Spanish, due to the arrival of other immigrants, keeps Los Angeles Spanish from changing entirely. She concludes by saying that the Spanish copula opposition with attributes is lost to a large extent among speakers in Groups II and III, but that these speakers are unlikely to pass the Spanish language to their descendants. These claims have been challenged by Gutiérrez (1992 & 1994) because he found that monolingual Mexicans from the same social class as those studied by Silva-Corvalán (1986 & 1994b) showed the same behavior regarding copula choice. Gutiérrez (1992& 1994) and Silva-Corvalán's (1986 & 1994b) findings evidence that further studies in language contact with regard to copula choice are needed.

Aguilar-Sánchez (2009, 2012& 2017) studied copula choice + *adjective* in the Spanish spoken in a multilingual region of Costa Rica called Limón. He studied the phenomenon in fifty-eight Labovian interviews of monolingual, bilingual,

and multilingual speakers aged 16 through to 91. His subjects had at least three years of formal instruction; twenty-six were males while thirty-two females, and thirty-nine reported non-restricted contact with English as opposed to nineteen with restrictive contact with English. Aguilar-Sánchez (2009, 2012& 2017) use a multi-feature model like the one used by Díaz-Campos and Geeslin (2005a & 2011) and Geeslin (2003), but he used a multi-level analysis of variance in the context of logistic regression, in which level one analyzes variance deriving from linguistic variables and level two analyzes variance from social variables. He set his significance value at 0.05, and his study had a power level of approximately 0.8. He found that linguistic features such as *experience with the referent*, *adverb* (presence or absence), *subject* (first or second order), *resultant state*, and *adjective class* are still variables that help to predict the use of one copula or the other. He points out that these findings are in tandem with previous studies of copula choice. In contrast with other studies on this phenomenon, Aguilar-Sánchez (2009, 2012 & 2017) found that social variables such as *age*, *education*, *bilingualism*, and *gender*also help predict copula use, when analyzed properly. He also found that the higher the level of contact with English formal instruction among bilinguals and multilinguals, the higher the use of *estar*; while the higher the level of contact with Spanish formal instruction, the higher the avoidance of the use of *estar* among monolinguals. He concludes that contact with English accelerates change in bilingual varieties of Spanish, giving further support to proponents of that hypothesis (e.g. Gutiérrez, 1992 & 1994; Silva-Corvalán, 1986 & 1994b). He also proposes that his findings suggest that two varieties of Spanish coexist in the same geographical area, and that these varieties are constrained by different social factors. Aguilar-Sánchez (2009, 2012, 2017) concludes that formal instruction aids language maintenance, but formal instruction in two languages also helps to accelerate change in contact varieties, and that the inclusion of social factors helps to explain this phenomenon beyond what linguistic factors can explain.

## Spanish Second Language Acquisition

This section deals with second language acquisition research regarding variation in copula choice and focuses on the relevant variables used in its multi-feature analysis. Geeslin (1999, 2000, 2002a, 2002b & 2003) has studied this phenomenon from several perspectives, providing new insights into the acquisition of *ser* and *estar* by non-native speakers. Because the nature of the present work is purely sociolinguistic, I will focus on elements linked to the study of copula choice by native speakers presented in Geeslin's line of work.

Geeslin (2003) examines contextual features predicting the use of *ser* and *estar* between native Spanish speakers and advanced Spanish learners. Thus, she examines whether the same features would explain the use of *ser* and *estar* in both groups. Geeslin (2003) found that semantic and pragmatic features interact in the prediction of copula choice. However, she found differences in the factors that determine copula use for each group.

Geeslin and Guijarro-Fuentes (2006) studied copula choice among native Portuguese speakers learning Spanish in contexts where the first and second language are similar, and where native speaker use varies. They found that *predicate type*, *susceptibility to change*, and *adjective* predicted the appearance of *estar*. Additionally, the *copulas allowed* (i.e., both copulas were allowed in the context) variable predicted copula selection for both native speaker groups. The model targeting the second language groups also showed an effect for *dependence on experience* (i.e., whether the speaker has first-hand or second-hand knowledge of the referent) and *animacy* (i.e., whether the referent was animate or not). These findings are attuned to all previous studies.

### Monolingual Spanish

Díaz-Campos and Geeslin (2004) applied the multi-feature model proposed in Geeslin's line of work (2000, 2002a, 2002b & 2003) to study copula choice in a variety of Spanish from Caracas, Venezuela, said to have limited or no contact with English. They used the Estudio Sociolingüístico de Caracas[4] (1987) [sponsored by the Consejo

---

4    Sociolinguistic Study of Caracas.

de Desarrollo Científico y Humanístico de la Universidad Central de Venezuela[5]] conducted in 1987 and 1988. The authors studied data from four age groups and three social classes (low, middle and upper). These groups were then evenly divided between men and women (2 participants per cell). The variables *predicate type, susceptibility to change, experience with the referent, resultant state, adjective, copulas allowed, age*, and *social class* were predictors of copula choice. They also found that older speakers tended to favor the use of *estar* and concluded that this pattern of behavior portrays the stability of this phenomenon in Caracas Spanish.

Díaz-Campos and Geeslin (2005a & 2011) investigated whether the extension of *estar* was a change in progress or a stable change. They researched how copula use differed from older and younger speakers by studying the frequency of use of *estar* and its predictors. They also studied how the frequency of occurrence of the predictors of *estar* would vary across age groups. In a reanalysis of the 2004 data, the authors found that older generations tend to favor the use of *estar,* and that there are no prominent differences between female and male speakers. In their predictor analysis by age group, they also found that *predicate type, resultant state, adjective*, and *copulas allowed* were common predictors for all age groups. However, some predictors were significant for specific age groups. These predictors were *susceptibility to change* for age group 14-29; *experience with the referent* for groups 14-29, 46-60, and 61 and above; and *socioeconomic class* for group 46-60. They concluded that a sociolinguistic interview does not elicit the same types of contexts from all speakers, and that three of the variables show a greater proportion of the categories that favor *estar* for older speakers. Their results on *age* and *gender* seem to indicate that the extension of *estar* is a stable phenomenon, but they also found that types of contexts were not equally represented among age groups, and that a comparison across age groups must also take these differences into account. Their results revealed that discourse factors such as *predicate type, resultant*

*state, adjective*, and *copulas allowed* are more important predictors than *age* and *gender*.

Geeslin and Díaz-Campos (2005) studied the extension of *estar* with adjectives and its relationship to language contact. They examined the individual categories of the *adjective class* variable to determine these categories' relationship with the use of *estar*. Their intention was to relate *adjective*s to the innovative use found in previous work (Geeslin, 2003). They stated that the best predicting classes of use of *estar* may not actually represent innovation. They were guided by questions such as what categories of the *adjective class* variable showed the most frequent use of *estar*, what categories of the *adjective* variable showed innovation, and how their results relate to previous studies on adjectives and copula choice. In a subsequent analysis of the data, they found that adjectives related to *mental state, size, physical appearance, color, status*, and *age* were more likely used with *estar*. The distribution of *estar* with each *adjective* across the variable *copulas allowed* showed more tokens of *estar* in the *both-allowed* category than were required. Their results revealed uses of *estar* in contexts formerly reserved for *ser* in all age groups. A high frequency of use of *estar* is especially noticed with adjectives of *age, size*, and *status*. Furthermore, when both copulas are allowed, speakers tended to favor *estar*, thus showing that innovative uses are emerging. They concluded that monolingual Caracas Spanish shows similarities to US/Mexican Spanish in its innovative use of *estar* with adjectives of age, *size, physical description, evaluation*, and *color* for certain age groups; adjectives of *status* and *size* for all age groups, and *age* and *description of personality* for the older speakers.

Aguilar-Sánchez (2007) studied the oral Spanish of 10 educated Costa Ricans and discovered that even though variables such as *adjective, predicate reading, experience with the referent*, and *resultant state* have been established as strong predictors of copula choice in Spanish as a Second Language and in monolingual varieties of Spanish (e.g., Venezuelan Spanish), the model is incomplete. Aguilar-Sánchez (2007) demonstrated that by including other variables related to the semantics of the adjective, some predictors of previous models

---

5    Scientific and Humanistic Development Council of the Universidad Central of Venezuela.

become non-significant. Variables included in his investigation provided evidence that the semantics of the adjective (i.e., *adjective class* and *resultant state*) play an equally prominent role as discourse/pragmatics features (i.e., *frame of reference*, *predicate reading*, and *experience with the referent*) in first language Spanish. Aguilar-Sánchez (2007) found that the inclusion of a semantic variable such as *gradiency* affects the selection of variables in the prediction of copula choice in educated monolingual Spanish. In fact, he demonstrated that a deeper look at this variable shed light on the fact that the choice between *ser* and *estar* in Spanish can be found in both discursive and semantic features. In his study, *ser* was used almost categorically to represent *individual-level* readings of the predicate while *estar* shows more variation in this area. He also found that *gradient* adjectives alternate with *absolute adjectives* with regard to predicate reading. He also found that *ser* is commonly used for *gradient* adjectives if the predicate has an *individual-level* reading, whereas *estar* is more common when the predicate is given a *stage-level* reading. However, he found that *estar* appears more frequently in *absolute* adjectives if the predicate is given an *individual-level* reading. Similarly, he found that *ser* appears more frequently if the predicate is given a *stage-level* reading. Aguilar-Sánchez (2007) concluded that the similarity of his results with those found in Díaz-Campos and Geeslin's (2005) analysis of this phenomenon by age group suggests that more studies are needed to support either hypothesis regarding copula choice in Spanish; change in progress (Silva-Corvalán, 1986) or stable change (Díaz-Campos & Geeslin, 2005a, 2011). Finally, evidence from Aguilar-Sánchez (2007) shows that ser the spread is not only of *estar* to the field once occupied exclusively by *ser*, but that the spread also happens in the opposite direction (i.e., *ser* is spreading to places that were once exclusive to *estar*).

New studies have explored copula choice in contact situations with languages other than English (Geeslin & Guijarro-Fuentes, 2007 & 2008; Guijarro-Fuentes & Geeslin, 2006), but because the scope of this paper is primarily related to contact of Spanish with English, they will not be discussed here in detail. These studies add to the body of knowledge on the study of copula choice.

## FREQUENCY EFFECT AND THE EXEMPLAR MODEL

This model proposes that language use shapes the form and function of the sound systems of a language (Bybee, 2001 & 2002). In this model, it is stated that language use creates and continually shapes language structure (Bybee, 2001 & 2002). Furthermore, the model contains gradient categories and relationships and is heavily affected by the nature of the input. It contains properties that can be subjected to formal modeling and simulation using connectionist architecture (Bybee, 2001 & 2002). In other words, it is possible to model the properties of the system through connectionist architecture to test it using empirical data.

Within this model, each token of a given lexical item experienced in use is stored as an exemplar of that item. This exemplar is accompanied by all its features in memory. While the storage takes place, the exemplar's phonetic and other details become part of its mental representations. As a result, instead of positing separately a lexicon and a system of abstract sound units that undergo transformation to surface forms by phonological rules; the Exemplar model favors the direct and detailed representation of words and frequent and infrequent phrases in memory. These representations form schemas which are organizational patterns in the lexicon and have no existence outside the lexical units from which they emerge (Bybee, 2001 & 2002). The frequency in which items occur is a crucial factor in determining how structure is affected. This functionalist approach to language analysis advocates that repetition is the mechanism that creates linguistic structure (Bybee & Scheibman, 1999). Furthermore, frequently occurring collocations may be processed and stored by the speaker as single units or chunks rather than as combinations of individual words. Thus, the more the speakers use it, the easier it is for them to access it, and the more fluent they become at using it.

### Change and the Exemplar model

Bybee (2001) states that…

[A] usage-based theory postulates that change is inherent in the nature of language. Grammars are not static entities, but constantly in the process of change resulting

from the way that language is used. Thus, a model of language change must include the mechanisms by which change occurs as an integral part of its architecture. (p. 57).

In her theory, Bybee (2001 & 2002) advocates that instead of accounting for sound change by the addition of new grammar rules, we need to see change as memory representations based on categorizations of tokens of use. Therefore, if the tokens begin to change, the core of the category also changes. The model accounts for the pattern of lexical diffusion and predicts that highly regular frequency items change before less frequent items. According to Bybee (2001 & 2002), this process of change involves at least three factors:

a. Sound change occurs as articulation proceeds. Change occurs in real time as language is used. Words and phrases that are used more often have more opportunities to undergo change.

b. Reduction also occurs within a discourse. Words that occur more often in the same discourse may also become more frequent words and undergo more reduction.

c. High-frequency words tend to be used in familiar social settings where there are fewer restrictions on reduction; therefore, they will undergo more reduction.

Under this model, change cannot be undone and has a permanent effect on the lexical representation of the words of a language. Because it is a model of phonology, it should be able to account for the present form of sounds and for the process by which phrases acquire their special phonological shapes. This model also assumes that the development of special phonological properties in high frequency words and phrases involves two properties of language. The first one is that high levels of use lead to reduction, and the second is that such reduction must be represented as part of their stored representation. Moreover, Bybee (2001, p. 32) states that the notion that frequency of co-occurrence in the input determines constituency has not been applied extensively at different levels of grammar. She concludes that the same principles apply for all levels of language. She proposes that linguistic objects (e.g., phonetic sequences, words, phrases)

are categorized the same way as non-linguistic objects (e.g., birds, chairs, noses). Therefore, this model can also be applied to explain changes in the syntax of languages.

### Spanish Copula and the Exemplar Model

Because the model accounting for the variation in copula choice is not complete (Aguilar-Sánchez, 2007, 2009, 2012 & 2017), I will focus on what has been established in the field and how this can be explained through a usage-based model. Most of the studies on copula choice have been done under the assumption that we need to look for the present elements of distinction in the latest developments of *estar,* and little attention has been paid to de Mello's (1979) claim that a clear understanding of the distinction between *ser* and *estar* is best achieved by considering them according to their functions: as main verbs, auxiliary verbs, or copulas. This that copulas are better achieved by considering all copula functions has also been made taking copula change as a unidirectional process: *estar* spreading to contexts of *ser* (Díaz-Campos & Geeslin, 2005a, 2005b, 2011; Silva-Corvalán, 1986, 1994a & 1994b); however, this change also seems to be going in the other direction where *ser* is spreading to contexts of *estar* (Aguilar-Sánchez, 2007, 2012 & 2017). More evidence is needed to prove the latter trend. The phenomenon in general is well explained by Geeslin (2002b) and her claim that the monolingual Spanish copula system is becoming a simplified one. Thus, what follows is a theoretical attempt to explain how copula choice is affected by use, and how we can account for variation in contexts where both copulas interact.

Following de Mello's (1979) call, I postulate that *ser* and *estar* have contexts in which they are not interacting; when they act as a verb in the case of *ser* (e.g. *ser* + Noun Phrase *es un hombre* "he is a man'") or when they act as an auxiliary in the case of *estar* (e.g. *estar* + Present Participle *está caminando* "s/he is walking"). Even though variation has not been found in these contexts, it is important to account for their frequency in people's speech, which helps us to explain a linguistic change through this model. Frequency of occurrence may explain why variation is not found in these contexts. Therefore, a usage-based model should consider how frequent each

copula is within the discourse and how the schemas are formed.

Variation, on the other hand, has been found in contexts where these two verbs function as copulas. These contexts include the following: *ser/estar + adjective*, *ser/estar + past participle*, *ser/estar + locatives*. Previous studies have shown a higher frequency of *ser* over *estar* in the data collected (Aguilar-Sánchez, 2007, 2012 & 2017; Díaz-Campos & Geeslin, 2004 & 2011).

However, very few studies have explored the effects of frequency of occurrence as possible explanations of the variation found in copula choice, and most of its theoretical approaches have been exploratory and lack consistency across studies. Thus, what I present here is an attempt to provide some consistency to the study of how frequency affects the prediction of copula choice in Spanish.

### Linguistic Context

Regardless of whether variation is found, a model to explain copula use in Spanish needs to consider all possible linguistic contexts in which copulas appear. This will help explain how speakers are using the copula and how these uses may help explain the direction of a specific change; in this case, the spreading of uses of *estar* to once exclusive contexts of *ser* or vice versa.

To start an analysis of natural data under the Exemplar model, it is important to propose a theoretical schema of the copula system in Spanish. This model should include all contexts in which both copulas are found, and it should reflect the findings of empirical studies regarding copula use.

Figure 1 shows possible schema formation with the copula function as the core and the other functions as peripheral components. In this figure, darker lines mean higher frequency of use. We can see how the category *Adjective* carries the heavier lines. Evidence of this comes from reports on percentages found in the empirical studies on copula choice mentioned above (Aguilar-Sánchez, 2007, 2009, 2012 & 2017; Díaz-Campos & Geeslin, 2004, 2005a, 2005b & 2011; Geeslin, 2003; Silva-Corvalán, 1986, 1994a & 1994b). This may be because the copula use of these verbs is heavily

**Figure 1.** Schema of the relationship of Spanish copulas
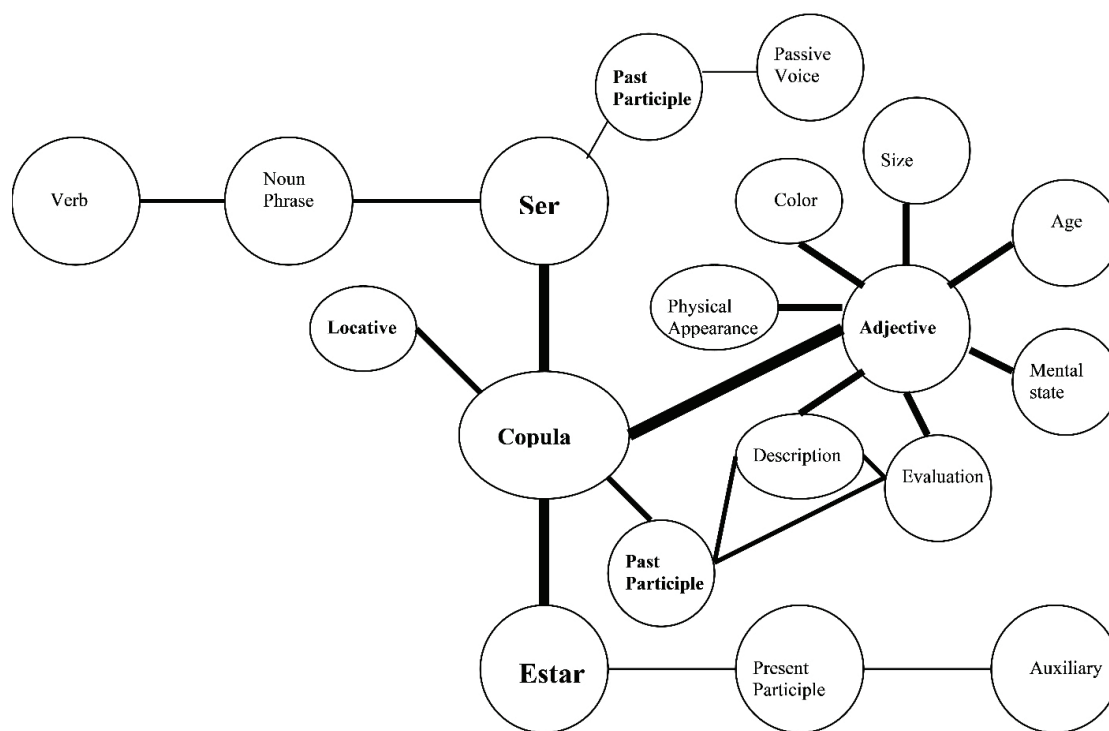


Source: Own elaboration

affected by their high frequency use with adjectives. All previous studies have focused on this particular context, and as pointed out in the previous sections, have found high levels of variation, which evidence that the system is changing. What is left to discover is how frequently these adjectives are used within the discourse in which they appear and within the language itself, and how its use affects or helps to the diffusion of the change.

In addition to the high frequency of the *copula + adjective* construction, we also need to consider how these verbs behave in the other contexts. It is possible that because attention has been paid mostly to the *copula + adjective* construction, some properties of the system might be left unnoticed. The model presented here attempts to explain any possible behavior of the Spanish Copulas regardless of its prominence in previous research.

## ACCOUNTING FOR FREQUENCY IN THE COPULA SYSTEM OF SPANISH USING EMPIRICAL DATA

To explain some phenomena in the Spanish copula system through the Exemplar model, I intend to suggest a procedure on how to analyze copula behavior in Spanish. This endeavor can be accomplished considering linguistic and social variables. These two types of variables will be analyzed at three levels using the latest advances in inferential statistics (multi-level analysis of variance) like the ones described by Gelman and Hill (2007) and Hox, Moerbeek, and van de Schoot (2017). These three levels are sources of statistical variance that affect the way the use of each copula is predicted statistically, and it comes from the use speakers give to them.

The first level includes the distribution of functions of each copula and their frequency. The second level would be the analysis of frequency and its effects on each of the components within each type of copula structures (e.g., within *copula + adjective*). The third level shows how social variables affect the use of each copula and their relationship to the frequency of use by function.

To account for frequency, one needs a way to account and measure frequency. For this, one can either build a frequency catalog of each word within the oral corpus being studied or collected or

use commercially available frequency dictionaries like Davies and Davies (2018) and Juilland and Chang-Rodriguez (1964) for Spanish and Davies and Gardner (2010) for English. The former is considered the preferred method, for it will yield frequency numbers related to the population being studied and it will also be based on oral data. The latter could be used to corroborate possible innovations in a particular variety of the language being studied, for example Limon's Spanish in Costa Rica, but because written language is highly edited, it may be detrimental to the study of actual use of the language in naturally occurring settings.

## THE STATISTICAL MODEL AND ANALYSES

### Level One Analysis

An analysis of data based on the Exemplar model for copula use in Spanish would consider an overall usage of *ser* and *estar* and their different structures and functions. Thus, one must first analyze the data for the appearance of all copula instances regardless of their accompanying elements. In this model, each copula instance is classified into *ser* or *estar*, our dependent variable. Then each instance is further classified according to the function it has in the context in which they appear. This classification creates some of our independent variables or predictors at this level. These functions include use as a *main verb*, an *auxiliary verb*, or a *copula*. Frequencies should be reported on each of these functions to be able to account for how frequency affects the behavior of each copula and to be able to establish any link between frequency and an explanation regarding the phenomenon under study, in this case the spread of *estar*.

The usage-based model predicts that highly frequent irregular forms tend to resist change (Bybee, 2001 & 2002). This can also help us to predict that adjectives that traditionally appear with *ser*, the most frequent form of the copula, and that are highly frequent will appear less with the other copula and vice versa. By accounting for the frequency of the functions of each copula, we can tell whether the regular form of *ser* and *estar*, regarding function, undergo change and whether highly frequent irregular forms of *ser* and *estar* resist it. Within this prediction, I hypothesized that highly

frequent contexts in which *ser* and *estar* are used categorically will tend to resist change, whereas low frequency contexts in which *ser* and *estar* are used categorically may show some sort of variation (i.e., change).

To test this hypothesis, one needs to include a variable that captures the frequency of use of these components in the model we wish to study. Therefore, each instance of copula should be coded as type of structure (i.e., copula, main verb, auxiliary). After coding for the type of structure they appear in, frequency counts should yield another categorization that can be divided into *low-frequency*, *medium-frequency*, and *high-frequency*. As stated above, frequency can be based on the appearance of each structure within the same corpus. Then we can use this frequency of use to determine, via inference, which function of these verbs is being used in innovative ways based on the dependent variables set forth as a construct, including frequency.

This first level of analysis can also be used as a qualitative analysis of the data when no inferential statistics are made regarding these different functions and structures because what we look for is how they are distributed in terms of usage. In other words, no predictions, only descriptions, are made. This qualitative analysis allows the researcher to see how Spanish copulas behave as a system in a purely descriptive way, which, in turn, is very important in the implementation of this type of model.

### Level Two Analysis

Once one has analyzed and understood how Spanish copula behaves in the corpus, one can proceed to analyze those contexts in which variation is present, for these contexts will help researchers analyze and explain the change under study. To conduct this second analysis, one of the functions should be selected. In this case, I would select the *copula + adjective*, for it is the one that shows more variation and has been studied the most. But, in multi-level analysis of variance, we can use this level 2 as a grouping level with random or fixed effects. By using this inferential model, we will be analyzing the entire system without isolating parts of it. Not isolating parts of the model allows us,

researchers, to find sources of variance that may be hidden when parts are taken in isolation.

As pointed out in the review of theoretical accounts and previous studies of copula choice in Spanish, several independent variables have *served* to predict copula choice in both first and second language Spanish. These variables are *adjective class*, *frame of reference*, *resultant state*, *predicate type*, and *gradiency*. Each of these independent variables contains at least two or more levels for classifying each token. I will not refer here to each of these levels; however, I will explain how to account for frequency, and how it can be used to explain the phenomenon under study.

Within the variable *adjective class* we find different types of adjectives; however, each of these types does not aid the prediction of copula choice in equal proportions. A possible explanation of this behavior is frequency of the adjectives that accompany each copula. Frequencies of each adjective should be regarded toto determine whether frequency plays any role in the selection of either copula. Therefore, a variable *adjective frequency* should be created, and each adjective can be classified as *high*, *mid*, or *low* frequency. The category *mid* is included tfor adjectives that present higher variation than those in the *low* category, but lower variation than those in the *high* category.

With respect to the variable *frame of reference*, we can also include a variable for frequency of the presence or absence of comparison. In this case, whether comparisons are more frequent than non-comparisons and how this may help us explain copula choice and the spread of *estar*.

*Resultant states* can also be categorized according to their frequency and a similar procedure as that followed for the variable *adjective class* should be used. In other words, a new variable called *resultant state frequency* should be created and each case of resultant states should be classified into *high*, *mid*, and *low* frequency within the corpus.

Similarly, *predicate type*, can be categorized according to frequency. In this case, each frequency refers to whether *stage-level readings* are more frequent than *individual-level readings* or vice versa, and whether it should be classified into *high* and

*low* frequency within the corpus as well. The results will help us further analyze the behavior of the copulas and test if the predictions presented by the Exemplar model apply.

*Gradiency* can be coded the same way. It is important to determine whether *gradient* adjectives are more frequent in our corpus than *absolute* adjectives or vice versa. This can also help us explain why some categories better predict the appearance of *estar* than others.

### Level 3 Analysis

After analyzing levels 1 and 2, one can include a level of analysis for social factors. This level includes all non-linguistic variables that form natural groups within the corpus. This level allows researchers to find sources of variation that can be caused by random use of the language, or patterns that are not random and used by the speakers according to social group. Frequency, in this case, can also be determined by looking at the usage speakers give to each of the functions of the verbs under study. So, for each subcategory of the variables at this level (e.g. *gender*, *level of education*, *dialect*, *bilingualism*, *age*, among others) frequency can represent how many uses of each function these groups produce to categorize them into *high*, *mid*, or *low frequency of use*. This categorization allows the researcher to infer whether different groups are using the copula in different ways like the ones discovered by Aguilar-Sánchez (2012), and whether these uses are innovative or not (i.e. producing change or not).

### CONCLUSION

After reviewing the theoretical accounts and previous studies on copula choice in Spanish and the properties of the Exemplar model, it can be concluded that a complete copula system study is needed to account for how frequently each of its components is used and how this affects the interpretation of our data. Thus, in order to understand whether a change is ongoing or stable, and whether the change is of *estar* extending to the contexts of *ser*, *ser* to the contexts of *estar*, or both, one needs to determine how each of the schema components interacts with each other, explain how these interactions fit into the traditional description of copula choice, move away from the generative view of language as a set of rules, and begin looking at it as a system highly affected by the frequency of use of its components (Bybee, 2001). One must also recognize the new advances in inferential statistics that may help us account for variance in ways we have not yet explored to understand a phenomenon like copula use. First, we need to see the system as a whole and study it as such t, then study its components and their function within this system in more detail. Frequency of use of the different parts of the model may help us to understand why researchers studying this phenomenon have found such varied results, even though it all points in one direction: Spanish copula use is undergoing change. Finally, we, as researchers, need to go back to the macro level to explain the change happening at the micro-level. Future research on copula choice should take the suggestions presented here for a more representative account of the copula system in Spanish. By achieving this representativeness, researchers may be able to approximate how the system works in specific social contexts and how it fits into the wider realm of the Spanish language.

### WORK CITED

Aguilar-Sánchez, J. (2007). The use of Spanish *Ser* and *Estar + adjectives*: A sociolinguistic pilot study on the oral Spanish of Costa Rica. 5th Hawaii Annual International Conference on the Arts and the Humanities, Hawaii.

Aguilar-Sánchez, J. (2009). *Syntactic Variation: The case of copula choice in the Spanish of Limón, Costa Rica* (Publication Number 3386661) [Doctoral dissertation, Indiana University-Bloomington]. ProQuest Dissertations and Theses Global.

Aguilar-Sánchez, J. (2012). Formal Instruction and Language Contact in Language Variation: the Case of *Ser* and *Estar + Adjective* in the Spanishes of Limón, Costa Rica. In K. L. Geeslin & M. Díaz-Campos (Eds.), *Selected Proceedings of the 14th Hispanic Linguistic Symposium* (pp. 9-25). Cascadilla Proceedings Project.

Aguilar-Sánchez, J. (2017). *Research Design Issues and Syntactic Variation: Spanish Copula Choice in Limón, Costa Rica*. Lambert Academic Publishing (LAP).

Andrade, M. J. (1919). The Distinction between *Ser* and *Estar*. *Hispania*, *2*(1), 19-23.

Bolinger, D. L. (1944). More on "*Ser*" and "*Estar*". *The Modern Language Journal*, *28*(3), 233-238.

Bybee, J. L. (2001). *Phonology and language use*. Cambridge University Press.

Bybee, J. L. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, *14*(3), 261-290. doi: https://doi.org/10.1017/S0954394502143018

Bybee, J. L., & Scheibman, J. (1999). The effect of usage on degrees of constituency: the reduction of don't in English. *Linguistics*, *37*(4), 575-596. doi: https://doi.org/10.1515/ling.37.4.575

Carlson, G. N. (1977). *Reference to kinds in English* [Unpublished doctoral dissertation]. University of Massachusetts-Amherst.

Clements, J. C. (2006). *Ser-estar* in the Predicate Adjective Construction. In J. C. Clements & J. Yoon (Eds.), *Functional Approaches to Spanish Syntax: Lexical Semantics, Discourse and Transitivity* (pp. 161-202). Palgrave Macmillan.

Davies, M., & Davies, K. (2018). *A frequency dictionary of Spanish: core vocabulary for learners* (Second edition. ed.). Routledge.

Davies, M., & Gardner, D. (2010). *A frequency dictionary of contemporary American English: word sketches, collocates, and thematic lists* (1st ed.). Routledge.

de Mello, G. (1979). The Semantic Values of *ser* and *estar*. *Hispania*, *62*(3), 338-341. doi: https://doi.org/10.2307/340597

Díaz-Campos, M., & Geeslin, K. L. (2004). *Copula Use in the Spanish of Venezuela: Social and Linguistic Sources of Variation* [Paper presentation]. New Ways of Analyzing Variation (NWAV), Ann Arbor, MI.

Díaz-Campos, M., & Geeslin, K. L. (2005a). *Is the extension of estar a change in progress or a stable change? A combined analysis of sociolinguistic and discourse factors in Spanish* [Paper presentation]. New Ways of Analyzing Variation (NWAV), New York.

Díaz-Campos, M., & Geeslin, K. L. (2005b). *The extension of estar with adjectives and its relationship to language contact* [Paper presentation]. Department of Spanish and Portuguese: Linguistic Brown Bag *Serie*s, Indiana University.

Díaz-Campos, M., & Geeslin, K. L. (2011). Copula use in the Spanish of Venezuela: Is the pattern indicative of stable variation or an ongoing change? *Spanish in Context*, *8*(1), 73-94. doi: https://doi.org/https://doi.org/10.1075/sic.8.1.04dia

Franco, F., & Steinmetz, D. (1983). *Ser* y *Estar* + adjetivo calificativo en español. *Hispania*, *66*(2), 176-184. doi: https://doi.org/10.2307/341394

Franco, F., & Steinmetz, D. (1986). Taming *ser* and *estar* with Predicate Adjectives. *Hispania*, *69*(2), 377-386. doi: https://doi.org/10.2307/341698

Geeslin, K. L. (1999). *The second language acquisition of copula choice in Spanish and its relationship to language change* [Unpublished doctoral dissertation]. The University of Arizona.

Geeslin, K. L. (2000). A New Approach to the Study of the SLA of Copula Choice. In R. P. Leow & C. Sanz (Eds.), *Spanish Applied Linguistics at the Turn of the Millennium* (pp. 50-66). Cascadilla Press.

Geeslin, K. L. (2002a). Semantic transparency as a predictor of copula choice in second-language acquisition. *Linguistics*, *40*, 439-468. doi: https://doi.org/10.1515/ling.2002.019

Geeslin, K. L. (2002b). The acquisition of Spanish copula choice and its relationship to language change. *Studies in Second Language Acquisition*, *24*, 419-450. doi: https://doi.org/10.1017/S0272263102003030

Geeslin, K. L. (2003). A Comparison of Copula Choice: Native Spanish Speakers and Advanced Learners. *Language Learning*, *53*(4), 703-764. doi: https://doi.org/10.1046/j.1467-9922.2003.00240.x

Geeslin, K. L., & Díaz-Campos, M. (2005). *The extension of estar with adjectives and its relationship to language contact* [Paper presentation]. Department of Spanish and Portuguese: Linguistic Brown Bag *Serie*s, Indiana University.

Geeslin, K. L., & Guijarro-Fuentes, P. (2006). Second Language Acquisition of Variable Structures in Spanish by Portuguese Speakers. *Language Learning*, *56*(1), 53-107. doi: https://doi.org/10.1111/j.0023-8333.2006.00342.x

Geeslin, K. L., & Guijarro-Fuentes, P. (2007). Linguistic and Social Predictors of Copula Use in Galician Spanish. In K. Potowski & R. Cameron (Eds.), *Spanish in Contact: Policy, Social and Linguistic Inquiries* (pp. 253-273). Benjamins.

Geeslin, K. L., & Guijarro-Fuentes, P. (2008). Variation in Contemporary Spanish: Linguistic Predictors of "*Estar*" in Four Cases of Language Contact. *Bilingualism: Language and Cognition*, *11*(3), 365-380.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Givón, T. (1984). *Syntax: A Functional-Typological Introduction, I*. Benjamins.

Givón, T. (2001). *Syntax: An Introduction, I-II*. Benjamins.

Guijarro-Fuentes, P., & Geeslin, K. L. (2006). Copula choice in the Spanish of Galicia: The effects of bilingualism on language use. *Spanish in Context*, *3*(1), 63-83. https://doi.org/10.1075/sic.3.1.06gui

Gutiérrez, M. J. (1992). The extension of *estar*: A linguistic change in progress in the Spanish of Morelia, Mexico. *Hispanic Linguistics*, *5*(1-2), 109-141.

Gutiérrez, M. J. (1994). *Ser y estar en el habla de Michoacán, México* [Unpublished doctoral dissertation]. Universidad Nacional Autónoma de México.

Hox, J. J., Moerbeek, M., & van de Schoot, R. (2017). *Multilevel Analysis: Techniques and Applications* (3rd ed.). Routledge. doi: https://doi.org/10.4324/9781315650982

Juilland, A. G., & Chang-Rodríguez, E. (1964). *Frequency dictionary of Spanish words*. Mouton.

Kratzer, A., Carlson, G. N., Pelletier, F. J., Krifka, M., Meulen, A. t., Link, G., & Chierchia, G. (1995). Stage-Level and Individual-Level Predicates. In *The Generic Book* (pp. 125-175). U of Chicago P.

Maienborn, C. (2005). A Discourse-Based Account of Spanish *ser/estar*. *Linguistics*, *43*(1), 155-180. doi: https://doi.org/10.1515/ling.2005.43.1.155

Merriam-Webster. (n.d.). Thesaurus. In Merriam-Webster. Retrieve from https://www.merriam-webster.com/thesaurus/thesaurus

Morley, S. G. (1925). Modern Uses of *Ser* and *Estar*. *PMLA/Publications of the Modern Language Association of America*, *40*(2), 450-489. doi: https://doi.org/10.2307/457234

Silva-Corvalán, C. (1986). Bilingualism and Language Change: The Extension of *estar* in Los Angeles Spanish. *Language: Journal of the Linguistic Society of America*, *62*(3), 587-608. doi: https://doi.org/10.2307/415479

Silva-Corvalán, C. (1994a). *Language contact and change: Spanish in Los Angeles*. Clarendon.

Silva-Corvalán, C. (1994b). The gradual loss of mood distinctions in Los Angeles Spanish. *Language Variation and Change*, *6*(3), 255-272. doi: https://doi.org/10.1017/S095439450000168X

Vendler, Z. (1957). Verbs and Times. *Philosophical Review*, *66*, 143-160.