

8

◆ The Sociolinguistic Profile of *Ser* and *Estar* in Cuban Spanish: An Analysis of Oral Speech

MANUEL DÍAZ-CAMPOS

Indiana University

IRAIDA GALARZA

Indiana University

GIBRAN DELGADO-DÍAZ

Indiana University

THE ANALYSIS OF COPULA choice in adjectival constructions has been the center of many investigations in Spanish, and they have focused on theoretical accounts (Clements 1988, 2006), sociolinguistic perspectives (e.g., Alfaraz 2012; Brown and Cortés-Torres 2012), first-language (e.g., Requena, Román-Hernández, and Miller, 2015), and second-language acquisitional contexts (e.g., Geeslin 2003), and language-contact situations in the US (Silva-Corvalán 1986) and Spanish in contact with other languages (Geeslin and Guijarro-Fuentes 2008). This work has contributed to our understanding of the factors involved in predicting copula choice as well as in describing the patterns of variation and change across varieties of Spanish.

The present investigation focuses on the extra-linguistic factors influencing copula choice in Cuban Spanish, a dialect of Spanish where recent research on this subject is scarce. This study also compares Cuban Spanish with other dialects such as Puerto Rican, Mexican, and Venezuelan Spanish in order to gain a broader understanding of the grammatical change involving [*ser/estar* + adjective] under a grammaticalization perspective. This chapter is organized as follows: First, a review of the previous literature is presented. In this section, a brief historical development of variation between *ser* and *estar* is discussed as well as analyses on the extra-linguistic factors that condition their use. The second section describes the present study and

the methodology used. In the results section, a statistical analysis of the factors conditioning the use of *ser* and *estar* in Cuban Spanish is offered and a comparative analysis across Latin American dialects is examined. The implications of these findings are discussed to gain a better understanding of the path of language change affecting [*ser/estar* + adjective] in contemporary Spanish. Finally, we offer some conclusions and future research directions.

Previous Literature

Traditional accounts have explained copular distribution in the construction [copula + adjective] in semantic terms, attributing different functions to each of the copulas, and thus attributing it a meaningful alternation. For instance, it has been argued that while *ser* expresses the permanent properties of a referent, *estar* expresses transitory, contingent, or circumstantial states (Gili Gaya 1961). For example, in 8.1 *ser* presents the quality attributed to the subject as permanent and in 8.2 *estar* presents the quality as contingent.

◆ Example 8.1.

Pedro es saludable.
Peter **is healthy.**

◆ Example 8.2.

Pedro está enfermo.
Peter **is sick.**

However, Gili Gaya (1961), based on the work of Hanssen (1913), explains that the distinction of *ser* and *estar* goes beyond the simplicity of the treatment it has received in the most traditional descriptive studies. Gili Gaya argues that in studying the distinction between *ser* and *estar*, the aspectual notion of imperfective and perfective as used by Hanssen may be useful. *Ser* would represent an imperfective predicate and *estar* would be used to indicate a perfective predicate. For instance, in example 8.3 *ser* is used to express a quality that is presented in its duration, while in example 8.4 the quality is seen as the result of an action, transformation, or change (both examples are from Gili Gaya 1961, 62).

◆ Example 8.3.

Este jarro es blanco.
This vase **is white.**

◆ Example 8.4.

Este jarro está roto.
This vase **is broken.**

Gili Gaya further explains that in an example like 8.1 the quality is seen without taking into account its origin, while in an example like 8.2, the quality is seen as the

result of a change that could be either real or assumed by the speaker. How do we know that a change has taken place according to this distinction? Gili Gaya argues that the experience of the speaker with the referent is a key factor. For example, in the case in 8.5 the use of *ser* + *fría* corresponds to a general description that is independent of the speaker's immediate experience. In contrast, example 8.6 represents an assessment of a description that depends on the immediate experience. (Both examples come from Gili Gaya 1961, 62.)

◆ Example 8.5.

La nieve es fría.

Snow **is cold**.

◆ Example 8.6.

Aquella nieve está fría.

That snow **feels cold**.

This subtle difference in the use of *ser* and *estar* in Spanish tends to be difficult for non-native speakers to capture. Gili Gaya explains that for English native speakers it is useful to consider if, in the sentence with a copula, one can use verbs such as to feel or to look (e.g., 'the coffee **feels hot**'/*el café está caliente*). In this case, not only is the perfective aspect present, the immediate experience with the referent is also seen. In the following section, a brief but concise review of the most important diachronic and synchronic patterns of change in copula choice is presented in order to provide the necessary context to understand current studies on this subject.

Diachronic and Synchronic Variation of ser and estar

Several scholars have argued that copular verbs in Spanish are in a process of diachronic change, in which *estar* is expanding to functions originally fulfilled by *ser* (e.g., Batllori and Roca 2011), as well as grammaticalizing new functions (cf. Torres Cacoullos 2011). *Ser* and *estar* have their roots in Latin *esse* and *stare*, respectively.¹ Latin *stare* had different meanings related to the posture and location of the referent, which can be translated as 'to stand,' 'to be situated,' and 'to stay' (Díaz-Campos and Geeslin 2011; Pountain 1982). As Batllori, Castillo, and Roca-Urgell (2009, 461) state, *stare* was used with a full lexical meaning in locative constructions in Latin, both in intransitive constructions (e.g., *Pugna stetit*, 'The battle continued') and constructions with locative expressions (e.g., *Stabat ad ianuam*, '[It] was at/in front of the door'). Pountain (1982) observes that *stare* was not nearly as frequent in either Classical or Vulgar Latin as it is now in contemporary Spanish. Thus, the locative function fulfilled by *stare* was also expressed by other Latin verbs (*adesse*, 'to be present,' 'to be here/there'; *sedere*, 'to sit'; and *esse*, 'to be,' among others). Moreover, Batllori, Castillo, and Roca-Urgell (2009) indicate that some copulative uses of *stare* were already present in Latin when the predicate was to be interpreted as "remaining or being in some place"; in other words, when the adjectival construction had a strong locative meaning. As a result, Batllori, Castillo, and Roca-Urgell

consider this construction with *stare* in Latin as the precursor of the copulative uses of *estar* in Spanish (see Batllori, Castillo, and Roca-Urgell 2009 and Batllori and Roca 2011 for further discussion). Consequently, the historic development of Spanish copulas, *ser* and *estar*, can be described as a process whereby *estar* has expanded its functional contexts of occurrence, overlapping with contexts in which *ser* was previously exclusively used (Batllori and Roca 2011; Marco and Marín 2015). Thus, the use of *estar* has developed from a locative construction in Latin, with a full postural meaning, to its current use in locative, attributive, and progressive constructions in contemporary Spanish (Batllori and Roca 2011).

It has been argued that the semantic distinctions between *ser* and *estar* in the attributive construction may have their origins in the copulas' diachronic lexical sources (Marco and Marín 2015). Consequently, the original locative meaning of *estar* has been extended to express temporally bounded states, such as temporally delimited adjectives (Marco and Marín 2015). This evolutionary path is consistent with semantic change of locative expressions. Bybee, Perkins, and Pagliuca (1994, 25) show that, crosslinguistically, many locative expressions develop temporal and aspectual meanings over time due to grammaticalization processes. Grammaticalization refers to a set of phonological, morphosyntactic, semantic, and pragmatic processes by which grammatical constructions are created out of discursive patterns of use (Bybee, Perkins, and Pagliuca 1994; Torres Cacoullous 2011, 149). These processes operate gradually in concrete instances of language use; thus, in this view, language use in context creates and shapes grammatical material (Bybee, Perkins, and Pagliuca 1994). For instance, the original locative meaning of *estar* developed different aspectual meanings both in [*estar* + gerund] constructions (which have a grammaticalized progressive meaning), and in [*estar* + adjective] constructions in Spanish. Bybee, Perkins, and Pagliuca (1994) argue that the temporal and aspectual senses that develop out of locative expressions have to do with the fact that both spatial and temporal meanings are present in locative expressions, even though spatial meaning is the principal sense of the form. Thus, the development of aspectual and temporal meaning consists of the loss of the spatial meaning and the strengthening of the temporal sense already present in the locative construction. In grammaticalization theory, the process whereby a form loses specific features of meaning is known as generalization of meaning, and is one of the many mechanisms involved in semantic change (Bybee 2010; Bybee, Perkins, and Pagliuca 1994). This type of semantic change has implications for the distribution of the grammaticalizing form since by having a more generalized or abstract meaning the form can occur in a greater number of contexts, thus extending its contexts of occurrence. This may result in instances in which two or more forms compete over a semantic domain, creating overlapping usage. Another possibility is that two forms develop contrasting meanings (e.g., was doing/did; Bybee, Perkins, and Pagliuca 1994). Regarding the grammaticalization path of copulative expressions, Devitt (1990) argues that there seems to be a crosslinguistic affinity between locatives and the expression of temporary states since many locative expressions across languages develop copulative functions with temporary states before gaining more general copulative functions. These findings reveal a pattern of grammaticalization according to which *estar* has

acquired new functions based on a process of semantic generalization, allowing it to occur in more copulative contexts.

Copula Choice in Cuban Spanish

Regarding previous research specifically focused on copula choice in Cuban Spanish, the only recent variationist investigation is presented by Alfaraz (2012). This scholar analyzed two sets of recorded interviews from Cubans. The first set belongs to *el Estudio Coordinado de la Norma Lingüística Culta de las Principales Ciudades de Iberoamérica y de la Península Ibérica* (Coordinated Study of the Educated Linguistic Norm of the Main Iberian-American Cities and Iberian Peninsula), which includes material recorded in the 1960s. The second set of recordings comprises sociolinguistic interviews of speakers who had recently arrived to the US from Cuba at the moment of the study (Cubans in Miami). This last set is from the 1990s. With respect to the variable context, Alfaraz did not include in the analysis contexts in which referents are compared to themselves in two different points in time. Therefore, this study did not examine the copular distribution taking into account the factor frame of reference, which was found to be significant in other studies (Brown and Cortés-Torres 2012; Díaz-Campos and Geeslin 2011; Gutiérrez 1994; Silva-Corvalán 1986). The analysis is limited as it only includes one linguistic factor, adjective type, disregarding crucial linguistic factors found to be significant in previous studies of copula choice (e.g., resultant state, frame of reference, susceptibility to change, and experience with the referent). The analysis also includes two sociolinguistic factors: generation and gender. The author reports 116 instances (19.3 percent) of “innovative” use of *estar*. ‘Innovative’ in this context is a categorization based on criteria established using reference grammars of Spanish, which do not necessarily reflect the actual corpus analyzed. Recall that recent research on other varieties of Spanish has revealed that *ser* is still the predominant form and that the contexts in which *estar* has expanded are somewhat more limited than what has been proposed in previous work. Regarding the effect of linguistic variable adjective type, Alfaraz (2012) shows that categories such as physical properties (e.g., *suave*, ‘soft,’ *sólido*, ‘solid’) favor the use of *estar* with a weight of 0.59 and evaluative adjectives (e.g., *bueno*, ‘good,’ *bello*, ‘beautiful’) also favor the use of *estar* with a weight of 0.56. However, while the use of *ser* and *estar* may be variable with these adjectives, it is not clear that complete neutralization of meaning distinctions has occurred with some of the adjectives included in these two categories. The analysis of social factors indicates that only the factor generation was selected as significant and included in the statistical model. These results show an increased use of *estar* by the younger generation of the 1990s, which, according to Alfaraz, may suggest a change in progress. However, the analysis of other relevant sociolinguistic factors may be needed to corroborate that interpretation. In summary, one of the limitations of this study is that it only takes into account a single linguistic factor, namely adjective class, to explain all copular alternation in Cuban Spanish, despite the evidence that copula choice is influenced by a composite of linguistic and social factors. Moreover, this study left out of the analysis contexts in which *ser* and *estar* can alternate, such as the

contexts in which the speakers compared referents to themselves. The next section is dedicated to describing the factors used in the present study.

Linguistic and Social Factors on Copula Variation

Several empirical studies have included a variety of linguistic and social factors in order to determine the nature of the variability of [*ser/estar* + adjective]. A crucial aspect in the design of these investigations has to do with the definition of the envelope of variation and the criteria used to classify *ser* and *estar* in attributive contexts. Related to the idea of the development of new contextual uses of *estar*, previous research has made the assumption that the synchronic variation in the [copula + adjective] construction observed in Spanish is due to the extension of *estar* into contexts previously occupied by *ser*. Several studies have focused on the description of the so-called innovative uses of *estar* (Alfaraz 2012; Silva-Corvalán 1986). Implicit in these studies is the definition of the dependent variable based on the categorization of cases according to “normative” uses of *ser* and *estar* as described in prescriptive Spanish grammars. In contrast, studies such as Brown and Cortés-Torres (2012) and Díaz-Campos and Geeslin (2011) have based their classification of the dependent variable in the actual cases found in the corpora by differentiating tokens with only *ser*, only *estar*, and both copulas, reflecting the distribution of them in the corpus. This methodology permits identifying the actual envelope of variation for copula choice according to speaker’s patterns of use. Table 8.1 summarizes all the linguistic factors that have been found to condition copula choice and the direction of the effects documented in these studies.

The Role of Social Factors

Processes of language variation and change can be mediated by social factors (e.g., Labov 1972). Social stratification of the use of *ser* and *estar* has been examined in several varieties of Spanish, including Venezuelan, Cuban, Puerto Rican, Costa Rican, and Mexican, among other dialects. However, in contrast to the effect of the linguistic variables discussed above, the results of the social variables are less consistent, revealing different patterns of variation according to the specific speech community under study. Table 8.2 summarizes all the social factors that have been found to condition copula choice and the direction of the effects documented in these studies.

These previous investigations and the independent variables discussed above are central pieces in the design of the present investigation. As will be described and explained in the methodology section, several factors included in the literature review are also used in the empirical analysis of Cuban Spanish in this study.

The Present Study

The present study provides an empirical, quantitative analysis of copula choice in contemporary Cuban Spanish using the tools of variationist sociolinguistics. The specific purpose is to examine the linguistic and extralinguistic factors predicting

◆ **Table 8.1.** Summary of linguistic factors previously studied according to the direction of the effect and investigations providing support

Factor	Direction of effect	Investigation providing support
Predicate type	Individual level → <i>ser</i> Stage level → <i>estar</i>	Aguilar-Sánchez 2009; Batllori, Castillo, and -Urgell 2009; Batllori and Roca 2011; Camacho 2012; Clements 1988, 2006; Díaz-Campos and Geeslin 2011, 2011; Geeslin 2003; Juárez-Cummings 2014; Leonetti 1994; Marco and Marín 2015
Resultant state	[- resultant] → <i>ser</i> [+ resultant] → <i>estar</i>	Aguilar-Sánchez 2009; Brown and Cortés-Torres 2012; Díaz-Campos and Geeslin 2011a; Geeslin 2003; Juárez-Cummings 2014
Semantic class of the adjective	Status → <i>ser</i> Observable traits → <i>ser</i> Mental and physical states → <i>estar</i>	Brown and Cortés-Torres 2012; Cortés-Torres 2004; Díaz-Campos and Geeslin 2011a; Gutiérrez 1992, 1994; Juárez-Cummings 2014; Ortiz López 2000; Silva-Corvalán 1986
Frame of reference	[- comparison] → <i>ser</i> [+ comparison] → <i>estar</i>	Brown and Cortés-Torres 2012; Díaz-Campos and Geeslin 2011a; Gutiérrez, 1994; Silva-Corvalán 1986
Susceptibility to change	Not changeable → <i>ser</i> Changeable → <i>estar</i>	Díaz-Campos and Geeslin 2011a; Juárez-Cummings 2014; Silva-Corvalán 1986
Experience with the referent	Indirect → <i>ser</i> Ongoing → <i>ser</i> Immediate → <i>estar</i>	Brown and Cortés-Torres 2012; Díaz-Campos and Geeslin 2011a; Guijarro-Fuentes and Geeslin 2006; Geeslin and Guijarro-Fuentes 2007 2008

the use of *ser* and *estar* in oral corpus from Havana, Cuba, since there are not recent studies that analyze the speech of Cubans on the island. This investigation also has the broad objective of describing this phenomenon in Cuban Spanish and comparing the results with other varieties, including Puerto Rican, Mexican, and Venezuelan Spanish, with the purpose of assessing synchronically an ongoing process of language variation and change. Following grammaticalization theory, a discussion of the findings is presented. With these objectives in mind the following research questions are proposed:

1. What are the linguistic and sociolinguistic factors that significantly predict copula choice in the [copula + adjective] construction in Cuban Spanish?
2. How does Cuban Spanish compare to other Spanish dialects (San Juan, Puerto Rico, Caracas, Venezuela, and Mexico City, Mexico) with respect to the linguistic and sociolinguistic constraints?
3. How can the results of the comparative analysis be accounted for within grammaticalization theory?

◆ **Table 8.2.** Summary of social factors previously studied according to the direction of the effect and investigations providing support

Social Factor	Dialect and direction of the effect	Investigation providing support
Socioeconomic status	Venezuela Lower class → <i>estar</i> Mexico Lower and middle class → <i>estar</i>	Díaz-Campos and Geeslin 2011; Juárez-Cummings 2014
Age	Puerto Rico 20–29 y/o → <i>estar</i> Mexico 35–44 y/o → <i>estar</i> Venezuela 46+ → <i>estar</i>	Brown and Cortés-Torres 2012; Díaz-Campos and Geeslin 2011a; Juárez-Cummings 2014
Level of education	Cuernavaca, México lower educational level → <i>estar</i> Costa Rica lower educational level → <i>estar</i>	Aguilar-Sánchez 2009; Cortés-Torres 2004
Gender	Costa Rica Women → <i>estar</i> Puerto Rico Men → <i>estar</i>	Aguilar-Sánchez 2009; Ortiz López 2000
Language contact	Bilinguals → <i>estar</i> Monolinguals → <i>estar</i> No uniform effect Basque and Galician bilinguals → <i>estar</i> Catalan and Valencian bilinguals → <i>ser</i> No effect	Brown and Cortés-Torres 2012; Geeslin and Guijarro-Fuentes 2008; Gutiérrez 1994; Ortiz López 2000; Silva-Corvalán 1986

Methodology

Participants

The corpus used for this analysis comes from a sample of native Spanish speakers from Havana, Cuba. The corpus is part of the Project for the Sociolinguistic Study of Spanish from Spain and America (Proyecto para el Estudio Sociolingüístico del Español de España y de América or PREESEA). The sample is composed of eighteen native Spanish speakers from Havana, Cuba. These recordings were recently completed and electronically published in 2014. The data contained in this sample are composed of sociolinguistic interviews lasting around 45 minutes for each speaker. The conversations included the following topics: greetings, weather, place where the subject lives, family and friends, traditions, danger of death situations, etc. The data is evenly distributed according to level of education (i.e., illiterate or primary education, high school, and college education), age (i.e., 20–34, 35–54, and 55 or more) and gender (i.e., male or female).

Envelope of Variation and Coding Scheme

The variable context was circumscribed following the notion of functional equivalence (Lavandera 1978), which delimits the envelope of variation in terms of a given grammatical or discursive function. Thus, it is not based on the traditional definition of sociolinguistic variables as “alternate ways of saying ‘the same’ thing” (Labov

1972, 188). As has been extensively discussed (Lavandera 1978; Schwenter 2011; Terkourafi 2011), the idea of equivalence in meaning is problematic in the case of linguistic variables that go beyond the phonological level, since the choice between two morphosyntactic structures can have an effect on the referential meaning expressed (Lavandera 1978).

In the present study, all cases of *ser* and *estar* that co-occurred with adjectives in the corpus were extracted. These [copula + adjective] constructions have a general attributive function in which a quality or characteristic is related to a referent (Díaz-Campos and Geeslin 2011; Geeslin and Gudmestad 2010). However, with this definition of the envelope of variation we are aware that not all of the cases examined have the same semantic or pragmatic interpretation, and in many cases the copulas are not necessarily interchangeable. Since the main goal of this study is to examine the variation between *ser* and *estar*, we left out other verbs that may also fulfill an attributive function, such as *sentir* (to feel) and *parecer* (to seem), but see Geeslin and Gudmestad (2010) for such an analysis. After extracting all of the relevant cases from the corpus, we coded each of them for five linguistic factors (*resultant state*, *frame of reference*, *dependence on experience*, *adjective class*, and *susceptibility to change*) and three sociolinguistic factors pertaining to the speaker (*age*, *education level*, and *gender*). The selection of factors was based on the previous literature and particularly following the methodology used in Díaz-Campos and Geeslin (2011). A description of the dependent variable and the independent variables are below.

Dependent variable: The dependent variable of this investigation is the copulative construction composed of [*ser/estar* + adjective]. As seems obvious from this description, the dependent variable has two variants:

- 8.6 *Ser*

...*más simples eeh tienen otro pensamiento ya cuando son **adultos** eeh su pensamiento va cambiando.*

...more simple ehh they have other ideas when they **are adults** eeh their ideas are changing. (LHAB_H13_073)

- 8.7 *Estar*

...*o la esperanza de que **esté más pintadita** / de que **esté más arreglada**.*

...or the hope that it **would be painted** / that it **would be in better shape**. (LHAB_M13_079)

Independent variables: Based on the review of the previous literature, this investigation tests the predictive power of five linguistic variables and three extralinguistic variables.² A description of these factors and examples of each one of them follow.

Resultant state: This variable used in previous studies (e.g., Aguilar-Sánchez 2009; Díaz-Campos and Geeslin 2011; Brown and Cortés-Torres 2012) distinguishes adjectives that derive from a dynamic situation (i.e., processes or activities and events or achievements and accomplishments; see a conceptual description in the work of Clements 2006). It is generally the case that these adjectives are derived from a verb form.

- 8.8 [+ Resultant]
*Son los lugares / que a esa hora **están abiertos** / que por lo menos a mí no me llaman mucho.*
They are places / that **are open** at that time / that at least they don't call me often. (LHAB_H23_085)

- 8.9 [- Resultant]
(23) *Mi apartamento **es muy amplio**.*
My apartment **is very big**. (LHAB_M13_079)

Semantic class of the adjective: As pointed out in Díaz-Campos and Geeslin (2011, 77) adjective class has received the widest attention in the previous literature. Some of the limitations of examining this variable are related to the fact that in naturally occurring data, some categories occur with greater frequency than others. For this reason, we have created broader categories to facilitate the quantitative analysis. The three categories created are mental states, observable traits (e.g., appearance, size, color), and adjectives of social class and status.

- 8.10 Mental states
*Personas de... de su generación / **están conscientes** de las necesidades de...*
People from... his generation / **are aware** of the needs of...
(LHAB_M23_091)
- 8.11 Observable traits
*Ya hablábamos de las características de él / él **es delgado** por naturaleza.*
We were talking about his characteristics / he **is thin** by nature.
(LHAB_M23_091)
- 8.12 Status
*...porque entonces ya el papá del niño y yo ya **estábamos separados**.*
...because by then the father of my son and I **were separated**.
(LHAB_M22_055)

Frame of reference: As explained above while reviewing the previous literature, this independent variable distinguishes between individual and class frames of reference in comparison contexts. An individual frame compares an entity to itself at different points in time. In contrast, a class frame compares a referent with a group of similar entities. It is expected that individual frame of reference will favor the use of *estar* based on empirical findings from previous research (e.g., Díaz-Campos and Geeslin 2011; Brown and Cortés-Torres 2012; Silva-Corvalán 1986).

- 8.13 [+ Comparison] Individual frame of reference
*No es un hombre feo // es trigueño bueno ahora **está canoso** porque ya **está viejo** // pero bueno.*
He is not ugly // he is tanned well now **he has grey hair** because **he is old** // but hey. (LHAB_M32_067)

- 8.14 [- Comparison] Class frame of reference
...porque mi mamá y mi tía son asmáticas y son alérgicas / fulminante al pelo de del gato.
 ...because my mother and aunt **are asthmatics** and **are allergic** /
 immediately to cat's hair. (LHAB_M11_007)

Susceptibility to change: This independent variable is designed to capture properties of the adjective and referent that can be considered changeable from those that cannot change (e.g., Díaz-Campos and Geeslin 2011; Juárez-Cummings 2014, Silva-Corvalán 1986).

- 8.15 Susceptible
Piden muchas hamburguesas / parece que / están muy buenas.
 They order lots of hamburgers / it seems that / **they are good**.
 (LHAB_H21_013)

- 8.16 Not susceptible
Pero mi sangre es gallega // es la realidad / y no lo puedo negar.
 My blood **is Galician** // it is the reality / and I cannot deny it.
 (LHAB_H33_097)

Experience with the referent: Originally proposed by Guitart (2002), this independent variable is design for the analysis of contexts where speakers indicate having firsthand experience with the referent. Three variants are distinguished for this variable: immediate experience, ongoing, and indirect experience.

- 8.17 Immediate
Me gusta pero como está hoy que hace sol sí está aceptable.
 I like it but how it is today that it is sunny it **is acceptable**.
 (LHAB_H12_037)
- 8.18 Ongoing
No fue una carrera muy fácil fue... fue bastante difícil / tuve que estudiar bastante.
 It was not an easy degree ... **it was** very **difficult** / I had to study hard.
 (LHAB_H13_073)
- 8.19 Indirect
...cosas diferentes / los amigos cuando / cuando uno es adulto por lo general se hace amigo de personas...
 ...different things / friends when / when one **is an adult** in general one becomes friends with people... (LHAB_M13_079)

The coding scheme also includes social independent variables for the *age*, *gender*, and *socioeconomic class* of the speaker. The next section provides some details concerning the statistical analysis used in the present investigation.

Analysis

The data were analyzed using Rbrul, a computer program designed for modeling sociolinguistic variation (Johnson 2009). The implementation of Rbrul is based on R statistical software. We performed a mixed-effects logistic regression statistical test. This type of analysis allowed us to examine the effects of the linguistic and social factors on copula choice in the [copula + adjective] construction, while accounting for the effect of the individual. In other words, the independent variables described above are treated as fixed effects, while the individual is treated as a random effect.

In the Rbrul analysis, the probability coefficients are expressed in log-odd units, which can be a positive or negative number: a positive log-odd value is interpreted as a favoring effect (i.e., an effect that favors the application value, such as *estar* in the present analysis), a negative value indicates a disfavoring effect, and 0 corresponds to a neutral effect (Johnson 2009). This is the natural logarithm of the odds of success, with the same concept as an odds ratio. Rbrul also expresses the probability coefficients in the traditional factor weight units, ranging from 0 to 1, so that the results obtained with the Rbrul analysis can be easily compared to studies that have reported their results using probabilistic weights (Johnson 2009). In the present analysis, we treated *susceptibility to change*, *frame of reference*, *dependence on experience*, *resultant state*, *adjective class*, *age*, *education level*, and *gender* as fixed factors, whereas *speaker* was treated as a random effect. The results of such analysis will be discussed in the following section.

Results

A total of 873 [copula + adjective] constructions were obtained from the corpus. *Estar* was used in 208 cases (23.8 percent), while *ser* had a higher usage rate in the [copula + adjective] construction, with 665 cases (76.2 percent). This copular distribution is consistent with other studies (e.g., Brown and Cortés-Torres 2012; Díaz-Campos and Geeslin 2011; Juárez-Cummings 2014), in which *ser* is the most frequently used copula or the default copula (Clements 2006, 182) in the adjectival construction. Clements (2006) attributes this distribution to the semantic underspecification of *ser*, which allows this copula to occur with a larger number of adjectives. To discuss the results of the regression, we follow the three courses of action commonly used in variationist sociolinguistics: (1) identify the significant factors that condition the variation; (2) examine the relative strength or significant contribution of each factor group (i.e., the magnitude of effect); and (3) identify the direction of the effect within a factor group (i.e., constraint hierarchy) (Poplack and Tagliamonte 2001, 92–94). These aspects of the results are considered in variationist research to represent the underlying variable grammar of a particular set of speakers.

Results from the mixed-effects logistic regression show that the best model for the prediction of [copula + adjective] in Cuban Spanish included the *speaker* as random effect, and *resultant state*, *frame of reference*, *dependence on experience*, and *adjective class* as fixed effects. The factors that did not significantly contribute to the use of *estar* were *susceptibility to change*, *age*, *education level*, and *gender*. Consequently, only semantic and pragmatic linguistic factors explained copula

choice in Cuban Spanish in the attributive function. These results are summarized in table 8.3.

Resultant state is by far the most significant predictor of *estar* in the Cuban data. An adjective classified as [+ resultant] favored the use of *estar*, with a weight of 0.862, whereas a [- resultant] adjective disfavored the use of *estar* with a weight of 0.138.³ The use of *estar* with [+ resultant] adjectives reflects the observation that there is an aspectual difference between *ser* and *estar* (Clements 2006; Fernández Leborans 1995; Luján 1981). In this case, this difference is manifested through the preference of *estar* with adjectives that have an underlying dynamic situation as source (Clements 2006). In other words, the adjectives that express the end point of a process or event tend to favor the use of *estar*. Thus, this result is consistent with the predictions made by Clements (2006) with respect to the different classes of adjectives across time stability categories, and their distribution with the copular verbs. This factor has also proven to be relevant in other dialectal areas such as Caracas, Venezuela, Mexico City, Mexico, and Puerto Rico (see comparative analysis below).

The second factor selected as significant by the model is a discourse-pragmatic one (Clements 2006; Geeslin 2003): frame of reference. Specifically, in contexts in which a referent was compared to itself [+ comparison], the use of *estar* was favored, with a weight of 0.844. In contrast, the cases in which the referent was classified as a member of a class disfavored the use of *estar* with a weight of 0.156. This is consistent with the results of other studies (Brown and Cortés-Torres 2012; Cortés-Torres 2004; Díaz-Campos and Geeslin 2011; Gutiérrez 1992, 1994; Juárez-Cummings 2014; Silva-Corvalán 1986). Moreover, the significant result of this factor provides empirical support to the semantic-pragmatic feature [+ nexus] proposed by Clements (1988), in which *estar* but not *ser* connects the referent to other (assumed or expected) situations or states. Consequently, these results show that *estar* tends to be favored in the discursive contexts when there is a comparison of the referent with itself.

The following significant factor included in the model was the discourse-pragmatic factor dependence on experience. An *immediate* experience with the referent favored the use of *estar* in the data, with a weight of 0.90, while both an *indirect* and an *ongoing* experience with the referent disfavored the use of *estar*. The preference for *estar* in contexts in which the speaker has a firsthand experience with the referent concurs with the results of other studies (Brown and Cortés-Torres 2012; Díaz-Campos and Geeslin 2011; Geeslin and Guijarro-Fuentes 2007, 2008; Juárez-Cummings 2014). The significant effect of this variable relates to the evidential (Roby 2009) and subjective uses of *estar* to express a reaction or surprise with respect to the subject referent (Guitart 2002).

Finally, adjective class also had an effect on the use of *estar* in the data. Particularly, mental adjectives favor the use of *estar*, with a weight of 0.786, whereas *status* and *observable traits* adjectives do not contribute to the use of *estar*. This result agrees with previous empirical studies (e.g., Brown and Cortés-Torres 2012; Díaz-Campos and Geeslin 2011; Juárez-Cummings 2014) and with theoretical accounts of the copular verbs in adjectival contexts (Clements 2006), since mental adjectives often represent transitory states.

◆ **Table 8.3.** Results of the mixed-effect model analysis with Rbrul indicating the linguistic factors that significantly favor [*estar* + adjective] in Cuban Spanish

Factors	Log-odds	Cases/ total	Percentage %	p-value	Centered Factor Weight
Resultant State				7.92e-41 ¹	
+ Resultant	1.834	96/131	73.3		0.862
– Resultant	–1.834	112/742	15.1		0.138
Range	72				
Frame of Reference				4.3e-42	
+ Comparison	1.691	99/149	66.4		0.844
– Comparison	–1.691	109/724	15.1		0.156
Range	69				
Dependence on Experience				8.46e-25	
Immediate	2.258	60/69	87		0.905
Indirect	–0.989	8/42	19		0.271
Ongoing	–1.269	140/762	18.4		0.219
Range	68				
Adjective Class				1.71e-07	
Mental	1.299	63/99	63.9		0.786
Status	–0.629	23/65	35.4		0.348
Observable traits	–0.670	122/709	17.2		0.339
Range	44				
Speaker standard dev					
	0.338				
Deviance	Df	Intercept	Mean	Input probability	
451.278	8	1.847	0.238	0.864	

1. P values are presented in scientific E notation. Therefore, $p = 7.92e-41$ equals 0.00000000000000000000792.

The model for copula choice in Cuban Spanish shows a rich patterning of both semantic (resultant state, adjective class) and discourse-pragmatic factors (frame of reference, dependence on experience). However, it is worth noticing that the predictor model did not include any social factors (i.e., age, level of education, gender). The lack of social stratification is consistent with other sociolinguistic studies that analyze morphosyntactic variation, in which the social factors do not necessarily have an effect in the observed patterns of variation. In fact, Díaz-Campos and Geeslin 2011 point out the following in their study of copula choice in Caracas Spanish: “While frequent and productive phonological change can be associated with certain social groups, syntactic change tends to be less frequent and its social value more elusive. In

fact, the history of copula choice in Spanish suggests a pattern of development with long periods of stability and the strong conditioning of linguistic factors” (2011, 91).

To further explore copula choice in Cuban Spanish, we analyzed the distribution of adjectives with each copular verb. Table 8.4 shows that *ser* occurs with a larger number of adjectives (241 different adjectives or 67.70 percent of the total adjective count), whereas *estar* occurs with a smaller number of adjectives (26.40 percent). It is worth noting that the adjectives that were used only with *estar* were mostly participial adjectives (59.57 percent of the total adjectives occurring with only *estar*), such as *restaurado* (restored), *sentado* (sat), and *vestido* (dressed), among others. When *ser* occurs with past participles it can express another function, that of passive voice (e.g., *Hasta de muerte fue herido*, ‘He **was** fatally wounded’ [LHABH21]). However, some participial adjectives are found with only *ser* in attributive contexts, such as *enamorado* (in love), *agradecido* (thankful), and *callado* (quiet), among others. These participial adjectives were used to describe personality traits of the referent (e.g., *Pero tenía un defecto, que era demasiado enamorado*, ‘But he had a defect, **he was** a womanizer’ [LHABH33]).

The adjectives that were used in the corpus with both *ser* and *estar* accounted for only 5.90 percent of the total adjectives. This number of variable adjectives was smaller than in the Caracas study (8 percent; Díaz-Campos and Geeslin 2011) and the Puerto Rico study (16 percent; Brown and Cortés-Torres 2012). Furthermore, similar to the results found by Díaz-Campos and Geeslin (2011), even though there were fewer variable adjectives, these were the most frequently used adjectives in the corpus, with an average rate of use of 11.33 times each. Table 8.4 illustrates the distribution of the nine most frequent variable adjectives in the corpus.

Regarding the variable adjectives, those shown in table 8.5 account for 70.59 percent of the total variable adjective cases. Furthermore, the majority of these adjectives belong to three adjective classes: evaluative class (57.90 percent of the total variable cases), such as *bueno* (good; e.g., *El cine europeo es muy bueno*, ‘European films are very good’ [LHABM33]; *El programa de television está bueno*, ‘The tv show is good’ [LHABH31]); age class (20 percent) such as *joven* (young) and *viejo* (old) (e.g., *cuando muchacho iba mucho... cuando era joven*, ‘when I was a teenager I used to go a lot... when I was young’ [LHABH23]; *me sentía bien, estaba joven*, ‘I felt good, I was young’ [LHABH32]); and physical appearance class (10 percent), such as *gordo* (fat; e.g., *físicamente me siento bien así porque estaba bien gorda*, ‘physically I feel well like this because I was very fat’ [LHABM21]; *me gustaba ser aeromoza pero era muy gorda*, ‘I liked being a flight attendant but I was very fat’ [LHABM2]). Taken together, these three adjectival classes account for 87.92 percent of the variable adjectives. Thus, variation is limited to certain adjective classes and the specific lexical items within them.

Tracking Synchronic Variation and Grammaticalization across Dialects

In this section, we compare the patterning of linguistic and sociolinguistic factors for copula choice in Cuban (Havana), Puerto Rican (Brown and Cortés-Torres 2012),

◆ Table 8.4. Distribution of *ser* and *estar* according to adjectives

	Number of adjectives	Percentage of adjectives	Number of cases	Percentage of cases	Average rate of use ¹
Only <i>ser</i>	241	67.70%	477	54.64%	1.98
Only <i>estar</i>	94	26.40%	158	18.10%	1.68
Variable: <i>ser</i> and <i>estar</i>	21	5.90%	238	27.26%	11.33
Total	356	100%	873	100%	

1. Following Díaz-Campos and Geeslin (2011b) and Geeslin (2013), we calculated the average rate of use of adjectives, the number of cases in each category was divided by the total number of adjectives in each category.

Venezuelan (Caracas; Díaz-Campos and Geeslin 2011), and Mexican (Mexico City; Juárez-Cummings 2014) Spanish in order to assess the synchronic grammaticalization stages of *ser* and *estar* in copulative constructions across these Spanish dialects. Recent investigations have proposed that grammaticalization processes can be observed synchronically in dialectal variation (Bybee 2010; Torres Cacoullós 2011; Silva-Corvalán 2001).

In order to carry out the comparisons across dialects, we follow the variationist comparative method (Poplack and Tagliamonte 2001). This method consists of comparing the independent regression models obtained for the different dialects, following the three lines of evidence mentioned above (i.e., significant factors, their relative strength, and the direction of their effects; Poplack and Tagliamonte 2001). Consequently, if the dialects under study share the same factor configuration along these three lines of evidence, and if the grammaticalizing forms show comparable frequency of use, it is taken as evidence that the dialects are in the same grammaticalization stage (Torres Cacoullós 2011). In contrast, dialectal differences may represent diverse grammaticalization stages or even dissimilar clines (Torres Cacoullós 2011). Therefore, comparing and contrasting the configuration of the linguistic factors that constrain the use of the copular verbs in different dialects may help determine the grammaticalization stage of the [copula + adjective] construction in Spanish.

In the present study, the cross-dialectal comparisons are possible because the studies chosen for the analysis used similar dependent and independent variables, and employed similar coding schemes. However, there are some differences across studies. For instance, Brown and Cortés-Torres (2012) did not include predicate type in their analysis, and split the adjective class factor into five categories (mental traits, physical traits, evaluation and description, age, and size), instead of the three categories used in the Cuban, Venezuelan, and Mexican studies (mental and physical traits, status traits, observable traits). Similarly, predicate type was not included in the Cuban analysis.

Another difference in the partition of the data has to do with the social factor age, since all of these studies classified their participants into different age groups. For instance, while the Mexican, Cuban, and Puerto Rican studies have similar age groupings (Mexican and Cuban age groups: 20–34, 35–54, 55+; Puerto Rican age groups: 20–29, 30–59 and 60+), in the Caracas study there were only two large

◆ Table 8.5. Distribution of *ser* and *estar* with the most frequent variable adjective in the Cuban corpus

Adjective	<i>Estar</i>	<i>Ser</i>	Total
<i>Bueno</i> (good)	7 (14.3%)	42 (85.7%)	49 (100%)
<i>Joven</i> (young)	4 (16.7%)	20 (83.3%)	24 (100%)
<i>Grande</i> (big)	4 (20%)	16 (80%)	20 (100%)
<i>Difícil</i> (difficult)	1 (6.3%)	15 (93.8%)	16 (100%)
<i>Igual</i> (same)	1 (7.7%)	12 (92.3%)	13 (100%)
<i>Viejo</i> (old)	9 (69.2%)	4 (30.8%)	13 (100%)
<i>Bonito</i> (pretty)	1 (9.1%)	10 (90.9%)	11 (100%)
<i>Chiquito</i> (small)	1 (9.1%)	10 (90.9%)	11 (100%)
<i>Fuerte</i> (strong)	2 (18.2%)	9 (81.8%)	11 (100%)

age groups that included much younger participants than the other studies (Caracas age groups: 14–45, 46+). Moreover, studies differed with respect to other extralinguistic factors: While the Mexican and Venezuelan studies included socioeconomic level as a factor (upper, middle, lower), the Cuban study analyzed the level of education of the speakers (illiterate, primary education, high school, or college education), and the Puerto Rican study coded for the degree of bilingualism of the participants (monolingual, bilingual) instead. Nonetheless, cross-dialectal comparisons can still be made since there are more similarities than differences in the coding schemes, especially with respect to the linguistic factors. The results of the four independent multivariate analyses are shown in table 8.6.⁴

Regarding the use of *ser* and *estar* in attributive constructions, table 8.6 shows that cross-dialectally *ser* is the predominant copula used in the majority of adjectival contexts (between 61 percent and 76.2 percent of the time, depending on the dialect). This result further corroborates the observation that *ser* is the default copula in Spanish (Clements 2006). Moreover, this distributional pattern suggests that *ser* has a more generalized meaning, which allows it to occur with a larger number of adjectives. For this reason, it is argued that the older Spanish copula, *ser*, has more copulative contexts of use due to its more grammaticalized (i.e., less specific) meaning. Similarly, the diachronic evidence put forward in the diachronic and synchronic section suggests that *estar* has also gone through loss of its specific locative meaning in the attributive construction, which has allowed it to expand its contexts of use within the copulative construction. However, the lower frequency of *estar* across dialects (between 23.8 percent and 39 percent depending on the dialect) in contrast to *ser* suggests that there are features of its meaning that restrict it to certain copulative contexts and adjectives. This idea will be explored next by taking into account the configuration of semantic and pragmatic factors across dialects.

The linguistic factors predicting the use of copulas in the dialects under study include the following independent variables in all varieties: resultant state, experience with the referent, and adjective class. All dialects studied (i.e., Cuban, Puerto Rican, Mexican, and Venezuelan) are conditioned in the selection of copula choice by these linguistic factors. While the impact of these independent variables may be

◆ Table 8.6. Linguistic and social factors that significantly favor the use of [*estar* + adjective] across Latin American dialects

Havana, Cuba (Present study)	Puerto Rico (Brown and Cortés-Torres 2012)	Mexico City (Juárez-Cummings 2014)	Caracas, Venezuela (Díaz-Campos and Geeslin 2011)
<i>ser</i> (76.2%)	<i>ser</i> (61%)	<i>ser</i> (62.43%)	<i>ser</i> (74.2%)
<i>estar</i> (23.8%)	<i>estar</i> (39%)	<i>estar</i> (37.57%)	<i>estar</i> (25.8%)
Resultant state	Frame of reference	Adjective class	Resultant state
+ Resultant 0.862	Individual-level 0.893	Mental/physical states 0.99	+ Resultant 0.92
– Resultant 0.132	Class 0.107	Observable traits 0.23	– Resultant 0.34
		Status 0.12	
Frame of reference	Experience with the referent	Experience with the referent	Adjective class
+ Comparison 0.844	Immediate 0.848	Immediate 0.95	Mental/physical states 0.88
– Comparison 0.156	Ongoing 0.321	Ongoing 0.47	Status 0.69
	Indirect 0.274	Indirect 0.09	Observable traits 0.42
Experience with the referent	Resultant state	Socioeconomic class	Predicate type
Immediate 0.905	+ Resultant 0.735	Lower class 0.74	Stage-level 0.79
Indirect 0.271	– Resultant 0.265	Middle class 0.67	Individual 0.37
Ongoing 0.219		Upper 0.06	
Adjective class	Adjective class	Resultant state	Experience with the referent
Mental states 0.786	Mental states 0.837	+ Resultant 0.88	Immediate 0.89
Status 0.348	Physical state 0.75	– Resultant 0.28	Indirect 0.58
Observable traits 0.339	Evaluation/description 0.392		Ongoing 0.48
	Age 0.313		
	Size 0.182		
	Age	Susceptibility to change	Susceptibility to change
	20–29 y/o 0.582	Changeable 0.87	Changeable 0.56
	30–59 y/o 0.505	Not changeable 0.31	Not changeable 0.33
	60+ 0.414		
		Age	Socioeconomic level
		35–44 y/o 0.64	Lower class 0.57
		20–34 y/o 0.51	Upper class 0.50
		55+ 0.39	Middle class 0.42
			Age
			46+ y/ o 0.57
			14–45 0.44
			Frame of reference
			[+ comparison] 0.59
			[– comparison] 0.48

different, the direction of the effect is the same across varieties. This is indicative that *estar* fulfills the same semantic and pragmatic functions across dialects: expressing resultant states comparing the referent with itself, describing immediate experiences with the referent, and, with predicates, describing mental states (changeable states and stage-level predicates). Frame of reference was selected as a significant linguistic predictor in three of the four dialects compared in the present study (i.e., Cuban, Puerto Rican, and Venezuelan). Susceptibility to change was found to be a significant linguistic predictor in dialects from Mexico and Venezuela. Finally, predicate type is only significant for Venezuelan Spanish. Recall that the direction of the effect in all cases shows the expected tendencies as acknowledged in table 8.1. The implication of these findings is that the dialects under study (Cuban, Puerto Rican, Mexican, and Venezuelan) seem to be in the same grammaticalization stage with respect to [copula + adjective] construction. Therefore, *estar* fulfills specific semantic and pragmatic functions within the copulative construction that are synchronically similar across Spanish dialects.

It is fair to say that social factors are at the periphery in the statistical models obtained in the different studies. This means that they can be considered the least significant factors, with some exceptions in the Mexican data. The factor age was selected as a significant predictor in Puerto Rican, Mexican, and Venezuela Spanish. Once again, while the magnitude of effect of this social variable is different according to the dialect, the direction of the effect in dialects from Puerto Rico and Mexico shows similarities with younger speakers favoring the use of *estar*. This does not necessarily suggest a change in progress as the present linguistic analysis indicates similar uses of *estar* across varieties. The fact that younger speakers used *estar* more often in some studies is not indicative of change in progress as there are some adjectives that are used only with *estar* and others, only with *ser*, while others occur with both copular verbs in a similar fashion. In addition, *estar* is used more by older speakers in Venezuela Spanish (see Díaz-Campos and Geeslin 2011 for a complete discussion of the social factors in this particular group of speakers). Furthermore, the fact that this variable was significant in Puerto Rico, Mexico, and Venezuela may be due to the nature of the interview, the adjectives used by the participants, or other outliers. Future research should take into account this issue in order to disentangle this possible effect. In the present study the social stratification of the copula choice is not evident as the statistical analysis suggests. Gender was not selected as significant in any of the studies, but socioeconomic class was selected as significant in the cases of Mexico and Venezuela, with lower socioeconomic groups favoring the use of *estar*. Therefore, in order to argue that there is a change in progress, future investigations have to account for the effect of the specific lexical items (i.e., individual adjectives) speakers use to better assess the influence of the sociolinguistic factors. This implies limiting the envelope of variation and the contexts in which neutralization has happened (see table 8.7).

Discussion

This investigation sought to analyze the linguistic and extralinguistic factors that constrain the use of *ser* and *estar* in copulative constructions in Cuban Spanish as well as

◆ Table 8.7. Summary of comparative analysis across dialects of Spanish

Factors	Cuba	Puerto Rico	Mexico	Venezuela
Linguistic Factors				
Resultant State	+	+	+	+
Frame of reference	+	+	–	+
Experience with the referent	+	+	+	+
Adjective class	+	+	+	+
Susceptibility to change	–	–	+	+
Predicate Type	n/a	n/a	–	+
Social Factors				
Age	–	+	+	+
Gender	–	–	–	–
Socioeconomic class	–	–	+	+

carry out a comparative analysis with other Spanish dialects. It was found that the use of *estar* in Cuban Spanish was restricted by semantic (resultant state, adjective class) and pragmatic (frame of reference, experience with the referent) factors. This finding concurs with recent investigations of these copular constructions because they share the same linguistic constraints and direction of the effect (see table 8.6). However, it was found that age, gender, and socioeconomic class were not significant. This result contrasts with some of the previous studies because Díaz-Campos and Geeslin (2011) and Juárez-Cummings (2014) found that socioeconomic class influenced the use of *estar*. Additionally, Brown and Cortés-Torres (2012), Díaz-Campos and Geeslin (2011), and Juárez-Cummings (2014) found that age significantly influenced the use of *estar*. Nonetheless, the findings of these previous investigations may be due to the nature of the interviews, the topics discussed, and the specific discursive contexts triggered by the interviews. Therefore, we recommend using an experimental task in order to control for these variables.

The second research question inquired about the dialectal comparison. This analysis revealed that *estar* is conditioned by the same linguistic factors. This allows us to determine, following Brown and Cortés-Torres (2012, 67), a cross-dialectal prototypical meaning of *estar* in attributive contexts: [temporally bounded, immediate, + comparative]. These results further dispute the claim made in several studies that there is a generational change in progress favoring *estar* (i.e., Alfaraz 2012; Silva-Corvalán 1986). On the contrary, we argue that the consistency of the linguistic constraints across dialects suggests a stable phenomenon in Spanish (Díaz-Campos and Geeslin 2011). However, this does not deny that the copulative structure in Spanish is slowly changing since grammaticalization processes operate gradually and are best observed in the diachronic dimension (Bybee, Perkins, and Pagliuca 1994; Torres Cacoullos 2011). According to Bybee (2010), the more we understand the diachronic development of structures the more we can explain their synchronic patterns of use.

Similarly, the results of this investigation suggest that the dialects under study (i.e., Cuban, Puerto Rican, Mexican, and Venezuelan) are in the same grammaticalization stage with respect to copulative uses. This hypothesis is based on the fact that the uses of *ser* and *estar* in attributive function have the same linguistic configurations (i.e., the same underlying variable grammar) across Spanish varieties. Even though there were differences in terms of the relative magnitudes of the effects of the significant factors, the direction of the effect (i.e., the constraint hierarchy) was constant across dialects.

This study also points out the importance of individual lexical items and their frequency of use on copula choice in Spanish. Regarding the results for Cuban Spanish, we found that a larger number of adjectives occurred with only *ser* in the corpus and that adjectives occurring with only *estar* were less numerous, while variable adjectives (i.e., those occurring with both *ser* and *estar*) were even scarcer in the data. However, variable adjectives had a higher rate of use in the corpus. Additionally, with respect to the adjective class, the multivariate analysis only selected mental adjectives as a significant predictor of *estar*, since the majority of mental adjectives (63 percent) were used with this copular form. However, our analysis shows that evaluative, age, and physical aspect adjective classes are also important to copula choice in Cuban Spanish, as they tended to show the most copular variability (i.e., they accounted for the 87.2 percent of the variable cases). Thus, it is argued that variation is mediated by certain adjective classes, the specific lexical items within these categories, and by their frequency of use.

Furthermore, grammaticalization theory can explain the patterns of use of *ser* and *estar* in copulative function in contemporary Spanish. According to Bybee, Perkins, and Pagliuca (1994), grammaticalizing forms interact with each other in the same functional domain by competing for the same uses or by covering different areas of meaning. For instance, there were cases in which the meaning of the constructions was subtle, as in examples 8.20 and 8.21. In these examples, the uses of *ser* and *estar* do not create a contrast in the interpretation, since both express a comparison of the referent with itself. On the other hand, there are contexts in which a contrast in meaning is maintained with the uses of *ser* and *estar*, such as in examples 8.22 and 8.23. In these examples, the adjective *buenas* in the construction with *estar* and a referent related to food (example 8.22) carry the interpretation of flavor and taste, whereas the same adjective in the construction with *ser* and a referent denotes behavior (example 8.23). Consequently, it is important to take into account the construction as a whole in the analysis, since, after all, grammaticalization processes operate within specific constructions and the particular elements within them have an impact on the resultant meaning (Bybee 2010; Torres Cacoullos 2011).

◆ Example 8.20.

Ahora es bella también pero en aquel momento era más bella.

Now she is pretty too but in the past she **was** even **prettier**. (LHABH22)

◆ Example 8.21.

*¡Qué **bella** está!*

‘How **beautiful you are!**’ (LHABH33)

◆ Example 8.22.

*Me piden muchas hamburguesas parece que **están** muy **buenas**.*

‘They order many hamburgers it seems that **they taste** really **good**.’
(LHABH21)

◆ Example 8.23.

*Mi niño **es bueno**, mi familia **es unida**.*

‘My child **is good**, my family **is close**.’ (LHABM23)

Conclusions

The main goals of this investigation were to determine the linguistic and extralinguistic factors that constrain the use of copula choice in Cuban Spanish, and to contrast its linguistic and extralinguistic configuration with Mexican, Puerto Rican, and Venezuelan Spanish with reference to the grammaticalization theory. The results of this investigation indicate that these dialects have similar linguistic constraints: resultant state, experience with the referent, and adjective class. However, it was found that these dialects differed in terms of the social factors that predict the use of *estar*. The present study did not find any social factor to be significant in Cuban Spanish. However, age was significant in Puerto Rican, Mexican, and Venezuelan Spanish and socioeconomic class was significant in Mexican and Venezuelan Spanish. We argue that these differences may be due to the specific discursive contexts of the interview. Therefore, it is suggested that these copular constructions are stable, contrary to what has been stated in previous investigations (Alfaraz 2012; Silva-Corvalán 1986). In fact, it is proposed that these constructions are at the same grammaticalization stage. Nonetheless, this does not reject the notion that the copulative forms in Spanish are slowly changing because grammaticalization processes operate gradually and are best observed in the diachronic dimension (Bybee, Perkins, and Pagliuca 1994; Torres Cacoullos 2011). Consequently, future studies should incorporate contact varieties of Spanish in order to complete a picture of the phenomenon within bilingual communities in comparison to monolingual varieties. Finally, there were instances where the distinctions between *ser* and *estar* were subtle, but there were also contexts where copulas maintained a clear contrast. These findings support the grammaticalization theory because grammaticalizing forms interact with each other in the same functional domain by competing for the same uses or by covering different areas of meaning (Bybee, Perkins, and Pagliuca 1994). On the other hand, this theory predicts that there are contexts in which a contrast in meaning is maintained. Therefore, it is argued that it is important to take into account the constructions and the surrounding elements in the analysis because grammaticalization processes operate within specific constructions and each particular elements have an impact on the meaning (Bybee 2010; Torres Cacoullos 2011).

Notes

The authors would like to thank Juan Escalona, the two anonymous reviewers, and the editor of this volume. Any mistakes remain our own.

1. Ford (1899) and Poutain (1982) mention that the precursor of Spanish *ser* was Vulgar Latin *essere* instead of Classical Latin *esse*. Moreover, Spanish *ser* obtained many of its forms from Latin *sedere* (to sit, to be seated), such as the infinitive form, the present participle, the future indicative, the conditional, the imperative, and the present subjunctive (Batllori and Roca 2011; Ford 1899).
2. Predicate type (individual-level and stage-level distinction) was not taken into account in the analysis of Cuban Spanish since it highly correlated with frame of reference factor ([+ comparison] and [– comparison] frames), resulting in the nonorthogonality (i.e., nonindependence) of the two linguistic factors. This correlation issue may apply to all Spanish varieties.
3. Since the dependent variable is binary in the present study, a favoring effect is interpreted as favoring the application value, *estar*, whereas a disfavoring effect can be interpreted as favoring *ser* (Tagliamonte 2012, 127).
4. The Cuban and Puerto Rican studies (Brown and Cortés-Torres 2012) performed a mixed-effects logistic regression through Rbrul, while the Mexican (Juárez-Cummings 2014) and Venezuelan studies (Díaz-Campos and Geeslin 2011) performed a binary logistic regression through Goldvarb X.

References

- Aguilar-Sánchez, J. 2009. "Syntactic Variation: The Case of Copula Choice in the Spanish of Limón, Costa Rica." PhD diss., Indiana University.
- Alfaraz, G. 2012. "The Status of the Extension of *estar* in Cuban Spanish." *Studies in Hispanic and Lusophone Linguistics* 5 (1): 3–27
- Batllori, M., E. Castillo, and F. Roca-Urgell. 2009. "Relation between Changes: The Location and Possessive Grammaticalization Path in Spanish." In *Diachronic Linguistics*, edited by J. Rafel, 443–93. Girona, Esp.: Universitat de Girona publication services.
- Batllori, M., and F. Roca. 2011. "Grammaticalization of *ser* and *estar* in Romance." In *Grammatical Change: Origins, Nature, Outcomes*, edited by D. Jonas, J. Whitman, and A. Garrett, 73–92. New York: Oxford University Press.
- Brown, E., and M. Cortés-Torres. 2012. "Syntactic and Pragmatic Usage of the [estar + Adjective] Construction in Puerto Rican Spanish: ¡Está brutal!" In *Selected Proceedings of the 14th Hispanic Linguistics Symposium*, edited by K. Geeslin and M. Díaz-Campos 61–74. Somerville, MA: Cascadilla Proceedings.
- Bybee, J. 2010. *Language, Use and Cognition*. Cambridge: Cambridge University Press.
- Bybee, J., R. Perkins, and W. Pagliuca. 1994. *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago: University of Chicago Press.
- Camacho, J. 2012. 22 "Ser and Estar: The Individual/Stage-Level Distinction and Aspectual Predication." *Handbook of Hispanic Linguistics* 69:453–74.
- Clements, J. C. 1988. "The Semantics and Pragmatics of the Spanish <COPULA + ADJECTIVE> Construction." *Linguistics* 26:779–822.
- . 2006. "Ser-estar in the Predicate Adjective Construction." In *Functional Approaches to Spanish Syntax: Lexical Semantics, Discourse, and Transitivity*, edited by J. C. Clements and J. Yoon, 161–202. London: Palgrave MacMillan.
- Cortés-Torres, M. 2004. "¿Ser o estar? La variación lingüística y social de estar más adjetivo en el español de Cuernavaca, Mexico." *Hispania* 87:788–95.
- Devitt, D. 1990. "The Diachronic Development of Semantics in Copulas." *Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistics Society* 16:103–15.
- Díaz-Campos, M., ed. 2011. *The Handbook of Hispanic Sociolinguistics*. Malden, MA: Wiley-Blackwell.

- Díaz-Campos, M., and K. Geeslin. 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.
- Fernández Leborans, M. J. 1999. "La predicación: Las oraciones copulativas." *Gramática descriptiva de la lengua española*, edited by I. Bosque and V. Demonte, 2354–460. Madrid: Espasa.
- Ford, J. D. M. 1899. "Sedere, Essere and Stare in the Poema del Cid." *Modern Language Notes* 14:4–10.
- Geeslin, Kimberly. 2003. "A Comparison of Copula Choice in Advanced and Native Spanish." *Language Learning* 53 (4): 703–64.
- . 2013. "Future Directions in the Acquisition of Variable Structures: The Role of Individual Lexical Items in Second Language Spanish." In *Selected Proceedings of the 15th Hispanic Linguistics Symposium*, edited by C. Howe, M. Lubbers, and S. Blackwell, 187–204. Somerville, MA: Cascadilla Press.
- Geeslin, K., and A. Gudmestad. 2010. "An Exploration of the Range and Frequency of Occurrence of Forms in Potentially-Variable Structures in Second Language Spanish." *Studies in Second Language Acquisition* 32 (3): 433–63.
- Geeslin, K., and P. Guijarro-Fuentes. 2007. "Linguistic and Social Predictors of Copula Use in Galician Spanish." In *Spanish in Contact: Policy, Social and Linguistic Inquiries*, edited by K. Potowski and R. Cameron, 253–73. Amsterdam: John Benjamins.
- Geeslin, K. L., and Guijarro-Fuentes, P. 2006. "Second Language Acquisition of Variable Structures in Spanish by Portuguese Speakers." *Language Learning* 56 (1): 53–107.
- . 2008. "Variation in Contemporary Spanish: Linguistic Predictors of *Estar* in Four Cases of Language Contact." *Bilingualism: Language and Cognition* 11 (3): 365–80.
- Gili Gaya, S. 1961. *Curso superior de sintaxis española*. Barcelona: Bibliograf.
- Guitart, J. 2002. "Spanish *ser* and *estar* in cognitive/pragmatic perspective." Paper presented at the Kentucky Foreign Language Conference, Lexington, Kentucky, April 19–21.
- Gutiérrez, M. 1992. "The Extension of *estar*: A Linguistic Change in Progress in the Spanish of Morelia, Mexico." *Hispanic Linguistics* 5:109–41.
- . 1994. "Simplification, Transfer, and Convergence in Chicano Spanish." *Bilingual Review* 19 (2): 111–21.
- Hanssen, F. 1913. *Gramática histórica de la lengua castellana*. Buenos Aires: Halle.
- Johnson, D. E. 2009. "Getting Off the GoldVarb Standard: Introducing Rbrul for Mixed Effects Variable Rule Analysis." *Language and Linguistics Compass* 3 (1): 359–83.
- Juárez-Cummings, E. 2014. "Tendencias de uso de *Ser* y *Estar* en la Ciudad de México." *IULC Working Papers*, 14 (2): 120–37.
- Labov, W. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Lavandera, B. 1978. "Where Does the Sociolinguistic Variable Stop?" *Language in Society* 7:171–82.
- Leonetti Jungl, M. 1994. "Ser y estar: estado de la cuestión." *Pliegos de la insula Barataria* 1:182–205.
- Luján, M. 1981. "The Spanish Copulas as Aspectual Indicators." *Lingua* 54:165–210.
- Marco, C., and R. Marín. 2015. "Origins and Development of Adjectival Passives in Spanish: A Corpus Study." In *New Perspective in the Study of ser and estar*, edited by I. Pérez-Jiménez, M. Leonetti, and S. Gumiel Molina, 239–66. Amsterdam: John Benjamins.
- Ortiz López, L. A. 2000. "Extensión de *estar* en contextos de *ser* en el español de Puerto Rico: ¿evolución interna o contacto de lenguas?" *Boletín de la Academia Puertorriqueña de la Lengua Española*, 98–118.
- Poplack, S., and S. Tagliamonte. 2001. *African American English in the Diaspora: Tense and Aspect*. Malden, MA: Blackwell Publishers.
- Pountain, C. 1982. "Essere/Stare as a Romance Phenomenon." *Studies in the Romance Verb: Essays Offered to Joe Cremona on the Occasion of His 60th Birthday*, edited by N. Vincent and M. Harris, 139–60. London: Croom Helm.
- Requena, P., A. Román-Hernández, and K. Miller. 2015. "Children's Knowledge of the Spanish Copulas *Ser* and *Estar* with Novel Adjectives." *Language Acquisition* 22 (2): 193–207.
- Roby, D. B. 2009. *Aspect and the Categorization of States: The Case of Ser and Estar in Spanish*. Philadelphia: John Benjamins.

- Schwenter, S. 2011. "Variationist Approaches to Spanish Morpho-Syntax: Internal and External Factors." In Díaz-Campos, *Handbook of Hispanic Sociolinguistics*, 123–47. Malden, MA: Wiley-Blackwell.
- Silva-Corvalán, C. 1986. "Bilingualism and Language Change: The Extension of *Estar* in Los Angeles Spanish." *Language* 62:587–608
- Terkourafi, M. 2011. "The Pragmatic Variable: Toward a Procedural Interpretation." *Language in Society* 40:343–72.
- Torres Cacoullós, R. 2011. "Variation and Grammaticalization." In Díaz-Campos, *Handbook of Hispanic Sociolinguistics*, 148–67. Malden, MA: Wiley-Blackwell.

For personal use, distribution prohibited. Author use with permission, (C) Georgetown University Press, 2017