The use of Spanish *Ser* and *Estar + Adjectives*: A sociolinguistic pilot study on the oral Spanish of Costa Rica

Jorge Aguilar-Sánchez: Indiana University
Introduction
The study of copula choice in Spanish has suggested that copula use is undergoing change.

- Part of this change is that the copula *estar* is becoming more accepted in an extended pre-adjectival context (Silva-Corvalán, 1986, 1994).
- This change has been attributed to contact with English; however, this has been challenged because monolingual Spanish also shows the same pattern (Guitiérrez, 1992, 1994).
This area of research has become a prolific one both in Spanish as a first language and in Spanish as a second language.

- These studies have been done in settings where
  - Spanish is a minority language (Silva-Corvalán, 1986, 1994 in Los Angeles Spanish)
- or in Spanish as a second language in settings where Spanish is a minority language (Geeslin, 2000, 2001, 2002 in Texas).
Motivation

- Even when it has been categorized as a monolingual society at large, Costa Rica is not different from most multilingual societies in the world.
  - “no nation is monolingual in the sense that all its citizens speak one and only one language, and probably no nation is monolingual in the less trival sense that everyone who lives there speaks the same language natively” (Thomason, 2004: 36).
The language situation in Costa Rica is more complex than what is normally believed.  

- Costa Rica, like most countries in Central America, is a country with native English speakers in a small percentage of its population (Aguilar-Sánchez, 2005a, 2005b) and native speakers of indigenous languages in a similar situation.
In this work I present previous work done in the study of Spanish copula, the results of the analysis of oral data, and a discussion of the findings.
Background
Costa Rica

- Located in Central America
- Languages (SIL, 2005)
  - Amerindian languages (6 alive and 1 extinct)
  - English (work-related diaspora)
  - Chinese (work-related diaspora)
  - Others (contemporary immigration)
Unlike English, Spanish has two copular verbs: Ser (to be) and Estar (to be)

Early descriptions of these verbs state that *ser* is durative and *estar* is transitory (Morely, 1925)

These two verbs cannot be described with absolute certainty because they have been found to be in variation (Silva-Corvalán, 1994; Gutiérrez, 1992, 1994).
Different approaches have been proposed to the study of copula choice. All these vary with regard to the definition and description of each copula. Despite of the differences in the definitions, these approaches have yielded a set of variables that have demonstrated to be strong predictors of copula choice.

These variables are:
Theoretical Accounts on Copula Choice

- Experience with the referent (subject)
  - This variable refers to whether the speaker has an *immediate, ongoing, or indirect* experience with the referent of the utterance (Andrade, 1919; Morely, 1925)

- Adjective class
  - This variable refers to the type of adjective (i.e., age, size, physical appearance, etc.) or the semantic making of it (Bull, 1942; Bolinger, 1944).
Theoretical Accounts on Copula Choice

- **Subject**
  - Bolinger (1944) proposes that the nature of the subject is an important feature on the choice of copula.
  - Clements (2006) sets forward that the type of subject is what influences copula choice in Spanish.

- **Susceptibility to Change**
  - This variable refers to whether the quality of the referent could change over time (Geeslin, 2003)

- **Frame of reference**
  - This variable establishes whether there is a comparison between the referent and him/herself at a different point in time or between him/herself and a group (Moellering, 1944; De Mello, 1979).
Theoretical Accounts on Copula Choice

- **Resultant state**
  - This variable distinguishes the use of copula + past participle as *passive voice* when used with *ser* and as *resultant state* when used with *estar* (*De Mello, 1979*).

- **Predicate reading**
  - This variable refers to whether the interpretation of the predicate is limited in time or not (i.e., whether the characteristic is permanent or transitory) (*Carlson, 1977; Clements, 2006; Kratzer, 1995; Maierborn, 2005*).
Theoretical Accounts on Copula Choice

- **Underlying structure**
  - This variable is related to the internal making of the adjective. It distinguishes between adjectives with underlying structures and adjectives with no underlying structures (Clements, 2006).

- **Gradiency**
  - This variable also refers to the internal making of the adjectives and distinguishes between adjectives that can be found in construction such as “more or less + adjective” (i.e., gradient) and adjectives that cannot be found in these constructions (i.e., absolute) (Givón, 1984).
Spanish in contact with English

Silva-Corvalán (1986, 1994)

- Explored the extension of *estar* in the speech of 27 bilinguals of different generations and degrees of Spanish language attrition.
- She found that *estar* is being used in contexts where *ser* is expected and that this phenomenon seems to be accelerated due to contact with English.
Previous work on the variation of copula choice

- **Monolingual Spanish**
  - Gutiérrez (1992, 1994)
    - Studied the oral speech of native Mexicans similar in characteristics to those studied by Silva-Corvalán (1986, 1994)
    - He found similar patterns to those found by Silva-Corvalán (1986, 1994) with respect to copula choice.
    - He concludes that the spread of *estar* to contexts of *ser* is not just due to contact with English.
    - Studied data from the Estudio Sociolinguistico de Caracas (1987)
    - Studied data from four age groups and three social classes, male and female and included 2 participants per cell.
    - They found that variables such as *predicate type, susceptibility to change, experience with the referent, resultant state, adjective class, copulas allowed, age, and social class* were strong predictors of copula choice.
Monolingual Spanish

Díaz-Campos & Geeslin (2005)

- Studied whether the extension of *estar* was a change in progress or a stable one.
- Did a reanalysis of data from the previous study.
- They found that older generations tend to favor the use of *estar* and there are no prominent differences between male and female speakers.
- They found that *predicate type, resultant state, adjective class* and *copulas allowed* were common predictors for all age groups.
- They also found that *susceptibility to change, experience with the referent*, and *social class* were predictors for different age groups. The latter two were predictors of copula choice in the higher age groups.
- They also found that older speakers tended to favor the use of *estar*. This pattern of behavior is an indication of the stability of this phenomenon in Caracas Spanish.
Previous work on the variation of copula choice

- Second Language Acquisition
  - Geeslin (2003)
    - Studied copula choice among 10 native speakers of Spanish and data from second language learners.
    - She found that responses to a contextualized questionnaire designed to identify those contexts in which only *ser* is allowed, where only *estar* is allowed, and context in which both copulas are allowed, were not homogeneous or categorical.
    - She proposes a different approach to the analysis copula choice in second language learner data.
The Present Study
The present study

- Methodology
  - Research questions
    - What linguistic and social variables help predict the use of *ser* and *estar*?
    - Are the variables the same for the prediction of *ser* as the ones for *estar* or are they different?
    - Is Spanish from highly educated Costa Ricans different from that of speakers of the same age range from other varieties of Spanish? More specifically, is there evidence of differences in copula choice?
The present study

- Participants
  - 10 oral interviews from the “Corpus de Habla Culta” (1987) of Costa Rica.
  - These speakers are university educated Costa Ricans who hold professional positions and represent the learned class of the country.
  - There are 5 men and 5 women
  - Their ages range between 28 and 50 years old
The present study

- Procedure
  - The data come from half hour Lobovian interviews conducted in 1987. These interviews included topics such as childhood memories, near-death experiences, and the present political situation of the country.

- Data Coding
  - Variables included in the analysis were those described above with the inclusion of gender as a non-linguistic variable.
  - All instances of (subject) + copula + adjective were accounted for and coded
The present study

Table 1
Coding of linguistic variables for the Study of Copula Choice

<table>
<thead>
<tr>
<th>Variable</th>
<th>Categories</th>
<th>Example</th>
<th>Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicate reading</td>
<td>[+ Stage-level]</td>
<td>Hoy, Elena está enferma ‘Today, Elena is sick’</td>
<td>Is the interpretation limited in time?</td>
</tr>
<tr>
<td></td>
<td>[- Stage-level]</td>
<td>Elena es simpática ‘Elena is nice’</td>
<td></td>
</tr>
<tr>
<td>Susceptible to change</td>
<td>[+ Susceptible]</td>
<td>El niño es pequeño ‘The boy is small’</td>
<td>Can the quality of the referent change?</td>
</tr>
<tr>
<td></td>
<td>[- Susceptible]</td>
<td>El coche es pequeño ‘The car is small’</td>
<td></td>
</tr>
<tr>
<td>Frame of reference</td>
<td>[+ Comparison]</td>
<td>El niño está alto ‘The boy is (grew) tall’</td>
<td>Is a comparison of the referent implied?</td>
</tr>
<tr>
<td></td>
<td>[- Comparison]</td>
<td>El niño es listo ‘The boy is smart’</td>
<td></td>
</tr>
<tr>
<td>Experience with the referent</td>
<td>[Indirect]</td>
<td>En España las fiestas son importantes ‘In Spain, festivals are important’</td>
<td>Does the speaker have first-hand knowledge?</td>
</tr>
<tr>
<td></td>
<td>[Ongoing]</td>
<td>Mi amigo Juan es desagradable ‘My friend Juan is unpleasant’</td>
<td>Is it on-going or an immediate reaction?</td>
</tr>
<tr>
<td></td>
<td>[Immediate]</td>
<td>Esta sopa está muy buena ‘This soup tastes great’</td>
<td></td>
</tr>
<tr>
<td>Resultant state</td>
<td>[+ Resultant]</td>
<td>El vaso está lleno ‘The glass is full’</td>
<td>Is the adjective a resultant state?</td>
</tr>
<tr>
<td></td>
<td>[- Resultant]</td>
<td>El vaso es caro ‘The glass is expensive’</td>
<td></td>
</tr>
<tr>
<td>Adjective class</td>
<td>[Age]</td>
<td>Joven ‘young’</td>
<td>Which semantic class best describes the adjective (in the sense it is used in the given context)? [note: both ‘description’ categories are last resorts]</td>
</tr>
<tr>
<td></td>
<td>[Size]</td>
<td>Grande ‘large’</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Physical appearance]</td>
<td>Gordo ‘fat’</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Description / evaluation]</td>
<td>Difícil ‘good’</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Description of a person(ality)]</td>
<td>Inteligente ‘intelligent’</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Color]</td>
<td>Azul ‘blue’</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Mental / physical state]</td>
<td>Animado ‘animated’</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Sensory characteristic]</td>
<td>Sabroso ‘tasty’</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Status]</td>
<td>Casado ‘married’</td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>[+first order]</td>
<td>Juan es joven ‘John is young’</td>
<td>Does the subject referent exist?</td>
</tr>
<tr>
<td></td>
<td>[- first order]</td>
<td>La situación está mala ‘The situation is bad’</td>
<td></td>
</tr>
<tr>
<td>Underlying structure</td>
<td>[Unidirectional Process]</td>
<td>Juan es alto</td>
<td>Does the adjective have an underlying dynamic situation? What type?</td>
</tr>
<tr>
<td></td>
<td>[Bidirectional Process]</td>
<td>Juan está mojado</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Unidirectional Event]</td>
<td>Juan es soltero</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Bidirectional Process]</td>
<td>Juan está levantado</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[N/A]</td>
<td>El hombre es mortal</td>
<td></td>
</tr>
<tr>
<td>Gradiency</td>
<td>[+gradient]</td>
<td>Juan está mojado</td>
<td>Can the adjective be seen in a “more or less” gradient?</td>
</tr>
<tr>
<td></td>
<td>[-gradient]</td>
<td>El hombre es mortal</td>
<td></td>
</tr>
</tbody>
</table>
The present study

- **Data Analysis**
  - The data were processed utilizing a VARBRUL analysis through Gold Varb X (Sankoff, Tagliamonte, & Smith, 2005) for Windows.
  - This is a binomial regression analysis of the factors included in the study.
  - The results show the factors that predict the source of errors encounter in the corpus with their corresponding log-likelihood and significant level as well as their probability weights.

- **Results**
  - A total of 337 tokens were found and coded following the criteria in Table 1.
## Results

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
</table>
| *Estar* as the application value  
Results of the logistic regression* |

<table>
<thead>
<tr>
<th>Factor Group</th>
<th>Factor</th>
<th>n</th>
<th>%</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 2: Adjective class</td>
<td>Description/Evaluation</td>
<td>14</td>
<td>7</td>
<td>0.315</td>
</tr>
<tr>
<td></td>
<td>Status</td>
<td>32</td>
<td>68</td>
<td>0.771</td>
</tr>
<tr>
<td></td>
<td>Mental/physical state</td>
<td>16</td>
<td>64</td>
<td>0.931</td>
</tr>
<tr>
<td></td>
<td>Size</td>
<td>5</td>
<td>29</td>
<td>0.736</td>
</tr>
<tr>
<td></td>
<td>Description of a person(ality)</td>
<td>1</td>
<td>3</td>
<td>0.411</td>
</tr>
<tr>
<td></td>
<td>Physical appearance</td>
<td>5</td>
<td>29</td>
<td>0.820</td>
</tr>
<tr>
<td>Group 4: Gradiency</td>
<td>Gradient</td>
<td>38</td>
<td>17</td>
<td>0.398</td>
</tr>
<tr>
<td></td>
<td>Non-gradient</td>
<td>36</td>
<td>28</td>
<td>0.668</td>
</tr>
<tr>
<td>Group 5: Predicate reading</td>
<td>Stage-level</td>
<td>38</td>
<td>64</td>
<td>0.909</td>
</tr>
<tr>
<td></td>
<td>Individual</td>
<td>36</td>
<td>12</td>
<td>0.380</td>
</tr>
<tr>
<td>Group 8: Experience with the referent</td>
<td>Ongoing</td>
<td>39</td>
<td>17</td>
<td>0.463</td>
</tr>
<tr>
<td></td>
<td>Indirect</td>
<td>25</td>
<td>23</td>
<td>0.469</td>
</tr>
<tr>
<td></td>
<td>Immediate</td>
<td>10</td>
<td>90</td>
<td>0.985</td>
</tr>
<tr>
<td>Group 9: Resultant state</td>
<td>Non-resultant</td>
<td>31</td>
<td>11</td>
<td>0.389</td>
</tr>
<tr>
<td></td>
<td>Resultant</td>
<td>43</td>
<td>68</td>
<td>0.877</td>
</tr>
</tbody>
</table>

*Log Likelihood = -80.602  p= .014*
Results

Figure 1: Comparison of Ser and Estar by Adjective Class
Figure 2: Comparison of Ser and Estar by gradiency

Weight 0.5

Gradiency
Results

Figure 4: Comparison of Ser and Estar by *predicate reading*
Figure 5: Comparison of Ser and Estar by *experience with the referent*
Results

Figure 6: Comparison of Ser and Estar by resultant state
With regard to the first question, the variables that help predict the use of *ser* and *estar*, in Costa Rican Spanish, in the *copula+adjective* construction are adjective class, gradiency, predicate reading, experience with the referent, and resultant state. Furthermore, the results of this study show that, as demonstrated in previous studies, linguistic and semantic variables are the strongest predictors of copula choice in the structure *copula + adjective*.

Concerning linguistic variables, the results show the following patterns: adjectives of size, physical appearance, mental/physical state, and status favor the use of *estar* (e.g., *no es porque estén esteticamente lindas* ‘it is not because they are aesthetically pretty’ [m30]) whereas adjectives of description/evaluation and description of a person(ality) disfavor it (e.g., *el sentir de la mujer es importante* ‘women’s feelings are important’ [m34]).

Because of the unexpected results regarding gradiency and the fact that they contradict previous theoretical accounts (i.e., Givón, 1984), a cross-tabulation of the variable gradiency with each of the other independent variables was conducted to account for the behavior of this variable in the data.
Figure 3: Relationship between *gradicency* and *predicate reading*
To answer question two, the evidence reveals that the same variables that help predict the use of *ser* help predict the use of *estar* and that *ser* and *estar* are in complementary distribution.

The statistical analysis reveals that we have accurately defined the contexts where *ser* and *estar* are in variation since the same factors were found to be significant for both *ser* and *estar* as the application value.

The analysis also shows that all variables were independent from one another and that there is no interaction among them. This evidence supports Geeslin’s and Díaz-Campos’ (2005) claim that we are dealing with a stable change and not a change in progress.
Regarding our third research question some of the results in this study are similar to those of previous studies regarding copula+adjective in Spanish.

First, the variable adjective class pattern similarly to the results in Díaz-Campos’ and Geeslin (2005) group age 30-45, in that Spanish from Costa Rica shows a distribution of copula use where estar is favored with adjectives of mental state and status. However, in Costa Rican Spanish for the equivalent age group, size and physical appearance seem to pattern differently with estar being the preferred copula for these adjective types.

With regard to adjectives of description/evaluation and description of person(ality) Caracas Spanish and Costa Rican Spanish pattern similarly with ser being the preferred copula and estar showing a high frequency that approaches a similar pattern to that of ser for the description of person(ality). In all adjective types, we can see the presence of estar patterning similarly to ser and/or being the preferred copula.
The present study has demonstrated that by including other variables that are related to the semantics of the adjective, some predictors of previous models become non-significant.

Further studies are needed to test the predictability of the new variables introduced in the present study as well as to include participants of the different age groups and social classes to study Costa Rican Spanish. Because the empirical evidence in this area has shown the power of discursive variables as well as semantic variables, more statistical model comparisons are needed in order to find the model that best fit the data regarding copula choice in copula+adjective constructions.
Author’s Information

Jorge Aguilar-Sánchez
Department of Spanish and Portuguese/Department of Second Language Studies

Email: jaguilar@indiana.edu