

***ESTAR* WITH AGE ADJECTIVES. A MULTIFACTORIAL BAYESIAN APPROACH TO THE DISTRIBUTION OF COPULATIVE VERBS ACROSS SPANISH VARIETIES \***

Silvia Gumiel-Molina, Universidad de Alcalá de Henares, c/ Trinidad, 3, 28801 Alcalá de Henares (Madrid), [silvia.gumiel@uah.es](mailto:silvia.gumiel@uah.es), ORCID: <https://orcid.org/0000-0002-3918-0489><sup>1</sup>

Norberto Moreno-Quibén, Universidad de Alcalá de Henares, c/ Trinidad, 3, 28801 Alcalá de Henares (Madrid), [norberto.morenoquibe@uah.es](mailto:norberto.morenoquibe@uah.es), ORCID: <https://orcid.org/0000-0002-7524-7636><sup>2</sup>

Isabel Pérez-Jiménez, Universidad de Alcalá de Henares, c/ Trinidad, 3, 28801 Alcalá de Henares (Madrid), [isabel.perezj@uah.es](mailto:isabel.perezj@uah.es), ORCID: <https://orcid.org/0000-0002-3040-3906><sup>3</sup>

Abstract

Using a Bayesian regression framework, this study examines the distribution of the Spanish copulas *ser* and *estar* with age adjectives across dialects based on PRESEEA corpus data. Results reveal a geolectal hierarchy: Mexican, Central American, and Andean varieties show the most advanced innovative use of *estar*, strongly associated with *cuando*-clauses and increasing use of the Imperfect tense, suggesting ongoing grammaticalization. European and Rioplatense varieties remain conservative. The findings confirm earlier proposals in the literature but challenge accounts that attribute innovative *estar* uniformly to subjectivity or evaluative meaning. Objective uses with age adjectives indicate partial neutralization with *ser* and a broader reorganization of the

---

\* We would like to thank the members of the *Theoretical Linguistics (TeLing)* Research Group at the University of Alcalá, Leticia Desborde-Zamorano, Diego Gibanel-Faro and Elisabeth González-Ortega for their help preparing the first draft of this paper. This research has been partially funded by the grant *Evidencialidad, Perspectivización y Subjetivización en las Interficies de la Lengua - II / Evidentiality, Perspectivisation and Subjectivisation at the Interfaces of Language - II*. Ministerio de Ciencia, Innovación y Universidades de España. Ref. PID2023-148755NB-I00.

<sup>1</sup> Author contribution: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Writing—Original Draft Preparation, Writing—Review & Editing.

<sup>2</sup> Author contribution: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Writing—Original Draft Preparation, Writing—Review & Editing, Software, Supervision.

<sup>3</sup> Author contribution: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Writing—Original Draft Preparation, Writing—Review & Editing.

copular system. The study argues that the phenomenon called ‘innovative *estar*’ reflects multiple converging processes shaped by adjectival class in some cases, grammaticalization stage, and regional variation.

## Resumen

Este artículo examina la distribución dialectal de las cópulas *ser* y *estar* con adjetivos de edad mediante un modelo de regresión bayesiana basado en datos del corpus PRESEEA. Los resultados muestran que las variedades de México, Centroamérica y la región andina presentan el uso innovador más avanzado de *estar*, asociado especialmente a cláusulas con *cuando* y al uso concomitante con el pretérito imperfecto, lo que sugiere un proceso de gramaticalización en curso. Las variedades europea y rioplatense se clasifican como conservadoras respecto al uso de la cópula *estar* con los adjetivos de edad. Los hallazgos confirman propuestas previas, pero cuestionan las explicaciones de este uso innovador basadas exclusivamente en la subjetividad. Los usos objetivos con adjetivos de edad indican una neutralización parcial con *ser* y una reorganización del sistema copulativo, reflejando variación estable más que sustitución completa.

## Keywords

Copulas, *ser*, *estar*, age adjectives, Spanish, subjectivity, perspective, Bayesian regression

## Palabras clave

Cópulas, *ser*, *estar*, adjetivos de edad, español, subjetividad, perspectiva del hablante, regresión bayesiana

## 1. INTRODUCTION

One of the phenomena of syntactic variation within Spanish that has received considerable attention from sociolinguistics, dialectology and theoretical linguistics is the so-called innovative use of the copula *estar* (see Gumiel-Molina, Moreno-Quibén and Pérez-Jiménez 2020, 2023 and 2024 for an exhaustive compilation of the relevant literature on this topic). In this paper, we will focus on the innovative use of *estar* with adjectives expressing age: *joven*, *pequeño*, *chico* ‘young’, *viejo*, *grande* ‘old’. According

to RAE-ASALE (2009, 37.9t), age adjectives in standard Spanish combine with *ser* to define a person by their age: *Juan tiene 20 años, es joven* ‘Juan is 20 years old, he is young’. When combined with *estar*, the predication expresses that the subject manifests some prototypical property associated with age in a specific situation. Sentences like *Juan ya está viejo* (‘Juan is already old’) mean that Juan is physically weak, has aged appearance or cannot perform some specific action because of his age (he is old to run the marathon, for example). However, in many American Spanish varieties—and marginally in Andalusian Spanish—*estar* appears in contexts where *ser* would be used in standard Spanish, as in (1), (2) (all the examples are extracted from the Preseaa Corpus; as will be clarified below; E refers to Interviewer; I refers to Informant):

(1) E: ah pues hay una falta de responsabilidad

I: una falta de responsabilidad de los padres y una falta / (...) de responsabilizar / a los hijos (...) a los hijos hay que responsabilizarlos desde que **están chicos** [MONR\_M32\_094]

“E: Well, there is a lack of responsibility

I: A lack of responsibility on the part of parents and a lack of / (...) of making children responsible (...) Children must be responsible since they are<sub>ESTAR</sub> young”

(2) I: y se amanecía tremenda rumba / tres días / <sic> cuatros </sic> días / eso era muy bonito // y / y los niños bueno // lo que tú tan <palabra\_cortada/> **los niños que estaban pequeñitos** / bueno le hacían comparsas / los disfrazaban / se ponían a inventar // así cosas bonitas de tradiciones para sacar de la rutina a la gente [CARA\_M21\_019]

“I: and dawn broke with tremendous revelry / three days / <sic> four </sic> days / it was very beautiful // and / and the children, well // what you <word\_cut\_off/> the children who were<sub>ESTAR</sub> very young / well, they made parades / they dressed them up / they started to invent // beautiful things from traditions to get people out of their routine”

Different linguistic subdisciplines have analyzed the various facets of this phenomenon of syntactic variation. Dialectology analyzes the geographical distribution of the preference for *estar* over *ser* with age adjectives in connection with a process of diachronic change still in progress that impacts the reorganization of the copular system;

for variationist sociolinguistics, the phenomenon constitutes a paradigmatic case of variation dependent on linguistic and extralinguistic factors (age, educational level, register, type of text, language contact) and provides an opportunity to analyze how certain innovations become consolidated in specific speech communities (De Jonge 1993a,b, Gutiérrez 1992, Cortés-Torres 2004, Malaver 2009, Díaz-Campos and Geeslin 2011, Aguilar-Sánchez 2012, Alfaraz 2012, Brown and Cortés 2012, García Márkina 2013, among others). On the other hand, for theoretical linguistics, the innovative use of *estar* challenges the classic explanations that account for the distribution of the copular verbs in Spanish based on the opposition between imperfective/IL (*ser*) and perfective/SL (*estar*) predications or between the expression of properties (*ser*) vs. resulting states (*estar*) (aspectual proposals) or based on the idea of comparison between different individuals (*ser*) or between stages/counterparts of the same individual linked to different situations (*estar*) (proposals based on *modes of comparison*) (see Silvagni 2021 for an overview of these lines of explanation, see also Gumiel-Molina, Moreno-Quibén and Pérez-Jiménez 2015). None of these approaches satisfactorily predicts the appearance of *estar* in contexts where stages/counterparts of the subject linked to alternative situations are not being compared, such as (1), (2). Thus, explaining the innovative use of *estar* forces a reconsideration of the semantic and/or structural features of either the copula or the age adjectives, or the formulation of syntax-semantics-pragmatics interface operations operating in the innovative varieties (Escandell-Vidal & Leonetti 2016, Sánchez-Alonso 2018, Sánchez-Alonso et al. 2019, Moreno-Quibén 2022, Gumiel-Molina, Moreno-Quibén and Pérez-Jiménez 2023, 2024, Escribano 2024, among others).

In this context, this paper aims to provide an updated descriptive study of the current distribution of the innovative use of *estar* with age adjectives across Spanish dialects. Section 2 will present the properties of the innovative construction <*estar* + age adjective>; specifically, it will present the generalizations found in the literature regarding the different types of syntactic-semantic contexts in which the innovative construction is documented (chronological, neutral and typological contexts, in De Jonge's 1993a,b terms). It will also analyze how the different types of proposals that have been formalized to explain the existence of innovative *estar*-predications with other classes of adjectives (namely evaluative adjectives) can be extended to account for the innovative structure with age adjectives. Section 3 presents the database, built from the Preseca corpus, which is used in this paper (a) to determine the geographical distribution of the copulas *ser* and *estar* (standard and innovative) with age adjectives in the different varieties of Spanish,

(a comparison that has only been made before in Malaver 2009, to our knowledge) and (b) to analyze the syntactic-semantic contexts in which the innovative structure with *estar* appears. The results obtained and the statistical analysis of the data offered in section 4 will show that the frequency of occurrence of the innovative construction <*estar* + age adjective> varies across geolectal varieties of Spanish, and, crucially, that there is also variation regarding the syntactic-semantic contexts in which the innovative *estar*-construction occurs (as first stated in De Jonge's 1993a,b, and also analyzed in Malaver 2009). The structure with (innovative) *estar* is the most widespread grammatical possibility to ascribe an individual to an age class (typological context) in Mexican, Central American and Andean Spanish, (3); in other American varieties (Caribbean and Chilean Spanish), however, the construction with *estar* (vs. *ser*) is less frequent, and furthermore, the innovative construction retains some aspects of the basic semantics of the copula *estar* in standard Spanish (i.e. comparison between alternative situations, neutral contexts), (4); the alternation with *ser* is also possible, (5). Finally, the innovative structure, is absent in some areas, like European and Rioplatense Spanish.

(3) I: después nos / él dijo nos vamos a casar le dije ay yo no quiero casarme/ sí sí  
 (...) mi papá quiere que te cases / que nos casamos de / que tú de blanco y que  
 yo así de / pero por la iglesia dije yo no

E: mjm / es que

I: y yo no / yo no quise

E: pero es que **estabas muy chiquita** / todavía te faltaba vivir más cosas /  
 conocer otras cosas [MONR\_M21\_044]

“I: Then he said, “We're going to get married.” I said, “Oh, I don't want to get  
 married.” “Yes, yes (...) My dad wants you to get married. He wants us to get  
 married. He wants you in white and me like this. But in church, I said no.”

E: Mhm. That's because...

I: And I didn't... I didn't want to.

E: But you were very young. You still had more to experience. You had other  
 things to learn.”

(4) I: ¿viste? / ahora ¿qué pasa? / que ahora yo / a lo mejor eso estoy haciendo ahora  
 disfrutando / disfrutando lo que no disfrute cuando **estaba adolescente**.  
 [SCHI\_M21\_019]

“I: Did you see? / Now what's going on? / Now I / Maybe that's what I'm doing now, enjoying / enjoying what I didn't enjoy when I was a teenager.”

(5) Inf. A mi mujer la conocí... la conocí cuando **éramos niños**.

Enc. -Sí, sí.

Inf. - Sí... Precisamente, unos años antes de que yo hiciera mi primer concierto en Veracruz [MC, Tape Mx-83, man, 57 years old, taken from De Jonge 1993a, ex. (5)]

“Inf. I met my wife... I met her when we were children.

Int. -Yes, yes.

Inf. - Yes... Precisely, a few years before I did my first concert in Veracruz.”

Finally, the generalizations found regarding the innovative use of <*estar* + age adjectives> will be interpreted within the complete paradigm of the innovative use of *estar*, with is also possible with other natural lexical-syntactic classes of adjectives. The idea will be defended that, in order to understand this paradigm in its entirety, it is necessary to take into account both the lexical-syntactic properties of the different classes of adjectives (as claimed in Moreno-Quibén 2022, Gumiel-Molina, Moreno-Quibén and Pérez-Jiménez 2023, 2024), as well as the process of grammaticalization of *estar* which would be undergoing an expansion with respect to the syntactic contexts in which it can appear (as claimed in De Jonge 1993a,b, Piñango and Fuchs 2023, Escribano 2024; and also claimed by Pérez-Jiménez, Gumiel-Molina and Moreno-Quibén 2018 to account for the differences between the copular systems of Catalan and Spanish). The innovative use of *estar* should be understood as a phenomenon that results from the confluence of various factors.

## 2. THE INNOVATIVE USE OF *ESTAR*

As noted in the Introduction, innovative *estar* predications do not attribute a property to the subject by comparing stages/counterparts of the entity referred to. In the following examples, illustrating the innovative structure with aesthetic (6) and age (7) adjectives, no comparison is established between stages or counterparts of the subject of predication with regards to the property expressed by *guapo* ‘handsome’ or *joven* ‘young’. (6) presents a hypothetical context where an individual is ascribed to a class; (7) expresses the chronological age of the individual. In this sense, innovative structures are different from standard copular sentences with *estar*.

(6) yo mi mundo lo hacía muy cerradito (. . .)/o sea nunca andaba con alguien que no bailara o que no fuera músico/ o que no **estuviera muy guapo** o sea o que no fuera como de la moda [MEXI\_M12\_048]

"I kept my world very closed (. . .)/I mean, I never hung out with anyone who didn't dance or wasn't a musician/or wasESTAR not very good-looking, I mean, or was not fashionable"

(7) [Context: Talking about the changes that Medellín has undergone]

(. . .) pues o sea los años cuando **yo estaba joven** // por ahí de quince de dieciocho años nunca pensé ver esta ciudad como está ahora [MEDE\_H21\_002]

“( . . .) mmm well the years when I wasESTAR young // around of fifteen of eighteen years never thought.1.sg see this city as is now”

Gumiel-Molina, Moreno-Quibén and Pérez-Jiménez (2024) pointed out (Table 1) that, although different lexical-syntactic classes of adjectives are documented in the innovative structure, age adjectives are most prominently found in this construction, followed by evaluative adjectives.<sup>4</sup> As can be seen in Figure 1, these authors also show that the innovative use of *estar* with age and evaluative adjectives is found in all varieties of American Spanish, albeit with uneven distribution.

Adjectival Class	Age	Evaluative	Dimensional	Property	Non-Qualifying	Total
General use	6 (1.4%)	95 (44.4%)	12 (20%)	102 (70.8%)	0	215 (25.4%)
Innovative use	413 (98.6%)	119 (55.6%)	48 (80%)	42 (29.2%)	10 (100%)	632 (74.6%)
Total	419 (100%)	214 (100%)	60 (100%)	144 (100%)	10 (100%)	847 (100%)

Table 1. Total number of general/standard and innovative *estar*-sentences by adjectival class (from Gumiel-Molina, Moreno-Quibén and Pérez-Jiménez 2024, Table 1)

<sup>4</sup> The prominent presence of age adjectives in innovative constructions has been noted in Gutiérrez (1992), Cortés-Torres (2004), Díaz Campos and Geeslin (2011), Brown and Cortés-Torres (2012), Alfaraz (2012), García-Márkina (2013), a.o.

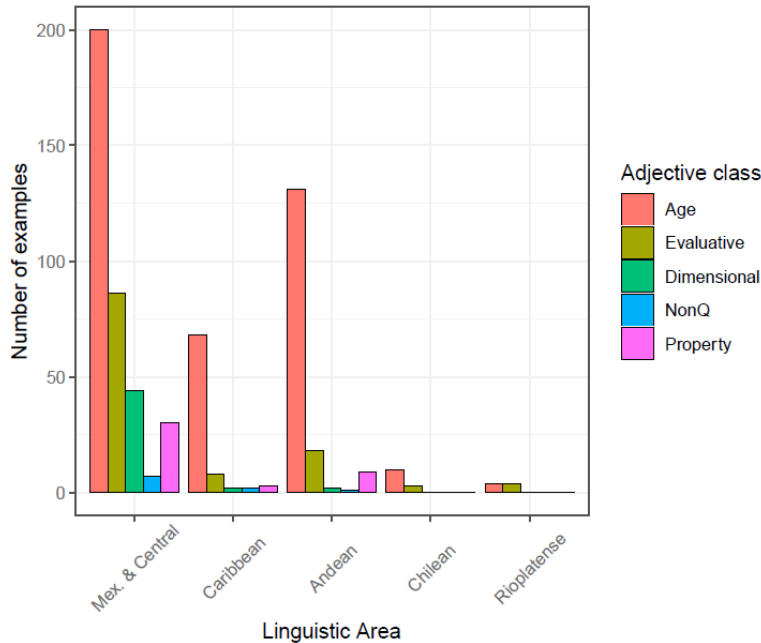


Figure 1. Number of examples of each adjective class in the innovative *estar*-structure by linguistic area. (from Gumiel-Molina, Moreno-Quibén and Pérez-Jiménez 2024, Figure 1). The adjective classes mentioned appear in the diagram from left to right within each area.

The first systematic generalizations about the innovative structure with *estar* and age adjectives appear in the works of De Jonge (1987, 1990, 1991a,b, 1993a,b). Analyzing oral corpora from Mexico City (Mexico) and Caracas (Venezuela), De Jonge (1993a,b) claims that innovative *<estar + age adjectives>* is documented with remarkable frequency even in formal registers in these areas, and that the phenomenon can't be attributed solely to social factors (as a case of free variation), but to the syntactic and pragmatic context of the predication.

The author bases his work on Falk's (1979) hypothesis to account for the *ser/estar* distinction in standard Spanish. According to Falk, standard *ser* (*Juan es gordo* 'John is fat', *Juan es viejo* 'Juan is old') is used to ascribe the subject to a category of individuals defined by a property (the subject of predication is thus identified by contrast with other individuals within a class, "class frame"). De Jonge calls these contexts in which the copula *ser* appears "typological contexts".

(8) [MC, p.88, woman, 55 years old, taken from De Jonge, 1993a, ex. (3)]

Ent. - ¿Y hay algún mínimo de edad para entrar, o desde qué edad...?

Inf. - Pues antes era de diez y ocho años. Ahora hemos pensado que **pueden ser más chicas.**

“Ent. - And is there a minimum age to enter, or from what age...?”

Inf. - Well, it used to be eighteen. Now we've decided that they can be<sub>SER</sub> younger”

Standard *estar* (*Juan está gordo* ‘John is fat’) ascribe the property to the subject of predication linked to a relevant situation in which the subject is involved (“individual frame”), which implies the possibility of change with respect to the property in an alternative situation. In (9) this situation is explicitly expressed by a functional standard (*muy A para X*, ‘too A for X’). De Jonge labels these contexts in which the standard copula *estar* appears as “chronological contexts.”

(9) [MC, Tape m-75, man, 25 years old, taken from De Jonge, 1993a, ex. (2)]

Inf. Siento que engordo, siento que... todo, me siento mal, pierdo condición [risas]... y eso que fumo bastante, o...

Ene. - ¿Entonces no te quieres preparar [sic] para las olimpiadas?

Inf. - No. [Risas] Y además, **yo estoy muy viejo para eso.**

“I feel like I'm getting fat, I feel like... everything, I feel bad, I'm losing my fitness [laughter]... and that's even though I smoke a lot, or...

Ene. - So you don't want to train for the Olympics?

Inf. - No. [Laughter] And besides, I'm<sub>ESTAR</sub> too old for that.”

Innovative *estar* examples are illustrated in (10), (11), (12). In this kind of examples, according to De Jonge, the communicative purpose of the utterance is not to classify the subject by their age with respect to other individuals (typological predication expressed by *ser* in standard Spanish), nor is the attribution of age to the subject linked to a potentially changing relevant situation (chronological predication, expressed by *estar* in standard Spanish). These are called “neutral contexts”. For example, in (12), the predication does not express the attribution of an age-property to a stage of the individuals in comparison or contrast with other stages, which would be the standard use of *estar*. In these contexts, *ser* is also possible, (recall (5)) (*ser* would be the standard option in these contexts).

(10) [MC, tape m-83, man, 57 years old, taken from De Jonge, 1993a, ex. (4)]

... Luis Horacio tuvo una decepción... una decepción muy grande, porque nosotros, **cuando estaba niño** lo llevamos al Conservatorio Nacional. Y hizo su solicitud y su, su prueba, y salió la prueba perfecta. Entonces, fuimos con el director por un... lo mandó llamar el director, el maestro Amparán. Y ése nos dijo que Luis Horacio no podía entrar, que porque tenía las manos chicas. “... Luis Horacio was disappointed... very disappointed, because when he was<sup>ESTAR</sup> young we took him to the National Conservatory. He submitted his application and took his audition, and he did perfectly. So we went to see the director... the director, Maestro Amparán, sent for us. And he told us that Luis Horacio couldn't be admitted because he had small hands.”

(11) [MC, Tape m-31, woman, 27 years old, taken from De Jonge, 1993a, ex. (1)]

... un viaje de uno o dos meses, entonces sería muchísimo tiempo de dejar a mis hijas, que pienso que **están muy pequeñas para dejarlas**. Tal vez **cuando estén un poquito más grandes**, me sienta con mayor libertad de actuar y puedo ir disfrazado realmente el paseo...

“... a trip lasting one or two months would mean leaving my daughters for a very long time, and I think they are<sup>ESTAR</sup> too young to be left behind. Perhaps when they are<sup>ESTAR</sup> a little older, I will feel freer to act and can really enjoy the trip...”

(12) I: [...] no me iría a vivir al sur <silencio/> porque<alargamiento/> (...) por

allá no es funcional vivir en el momento en que estamos <silencio/> demasiados trancones<alargamiento/> demasiada gente<alargamiento/> <vacilación/>hay demasiada gente<alargamiento/> demasiado conflicto / demasiadas cosas <silencio/> demasiado <vacilación/> hay / cuando hay demasiada concentración de personas <silencio/> aumentan los riesgos / haga de cuenta como entrar a un estadio / a ver algún partido o algo así <silencio/> yo recuerdo una vez / <vacilación/> llevé a mi hijo / **estaba pequeño** <silencio/> / porque a mí me llevaban **cuando estaba pequeño** / y<alargamiento/> y<alargamiento/> y para entrar // cuando el borbotón de gente / me tocó acomodar de una manera / que mi hijo quedara protegido / <vacilación/> no me gustan las concentraciones (...) [CALI\_H32\_023]

“I: [...] I wouldn't go to live in the south <silence/> because<prolongation/> (...) it's not functional to live in the moment we're in <silence/> too much traffic<prolongation/> too many people<prolongation/ > <hesitation/>there

are too many people<prolongation/> too much conflict / too many things  
 <silence/> too much <hesitation/> there is / when there is too much  
 concentration of people <silence/> the risks increase / think of it like going  
 into a stadium / to see a game or something like that <silence/> I remember  
 once / <hesitation/> I took my son / he was<sub>ESTAR</sub> little <silence/> / because  
 they took me when I was<sub>ESTAR</sub> little / and<elongation/> and<elongation/> and  
 to get in // when the crowd surged / I had to position myself in such a way /  
 that my son was protected / <hesitation/> I don't like crowds”

De Jonge claims that this extension of *estar* from (standard) chronological contexts to neutral contexts may have a pragmatic origin, based on the speakers' individual intention to attribute a property to a subject connected to a situation of particular relevance to themselves, which could also have changed or potentially change. Thus, in neutral contexts, the age-predication is interpreted as the temporal-frame of a (potentially changing) relevant situation, according to the speaker's perspective, hence retaining the basic contrast-of-situations semantics of *estar*, as is clearly observed in the previous examples. Therefore, according to De Jonge, neutral contexts are identified by expressions denoting temporal frame like *cuando*. The extension from the chronological context to the neutral context is favored by the concurrence of explicit syntactic markers typical of the standard chronological use, such as modifiers expressing temporal phase (*ya, todavía*), as De Jonge claims, and also, as Escandell-Vidal and Leonetti (2016) notes, by first-person subjects, referring to the speaker's perspective.

A further contextual extension of *estar* implies its use in the typological contexts, as shown in (2), (3). As De Jonge notes, *cuando* may be absent in typological contexts. Accordingly, we will assume that the verbal tense alone might become a syntactic clue for the grammaticalization of *estar* in the typological contexts; given that the interviewees recount past events in their lives, it is expected that the Imperfect tense will appear prominently (as in (2), (3) above), though other tenses are also possible, like the Present tense (13).

- (13) I: se chivea el puto se chivea es que no le gusta ¿sabes por qué no le gusta salir conmigo? / porque<alargamiento/> te digo que me lleva <vacilación/> casi veinte años / y él dice que le da pena que lo vean conmigo porque van a ver dicen que **ya está viejo** y son muchos ojetes me dicen <cita> el viej

<palabra\_cortada/> </cita> < cita> chinga tu madre **no está viejito** pendejo  
</cita> a mí me encabrona que le digan que **esté viejito** a si de veras **estuviera  
viejito** güey ya me hubiera buscado un querido [PUEB\_M21\_056]

“I: He's snitching, the bastard is snitching, he doesn't like it. Do you know why he doesn't like going out with me? / because<elongation/> I'm telling you, he's been taking me <hesitation/> for almost twenty years / and he says he's embarrassed to be seen with me because they're going to see him and say he's<sub>ESTAR</sub> old, and there are a lot of assholes who tell me <quote> the old man <word\_cut/> </ quote> <quote> fuck your mother, he's<sub>ESTAR</sub> not old, asshole </quote> it pisses me off when they say he's<sub>ESTAR</sub> old, if he really were<sub>ESTAR</sub> old, man, I would have found myself a lover already.”

According to De Jonge, from the synchronic point of view, those geographical areas where *estar* is found both in neutral contexts and typological contexts are considered more innovative than the areas where *estar* is only found in neutral contexts (De Jonge 1993a,b only consider Mexico City vs. Caracas, but Malaver 2009 extends the analysis to the major dialectal areas of Spanish, as will be described below). These contexts can be grammatically identified: temporal-frame markers like *cuando* can be considered synchronically indexes linked to neutral contexts (in concurrence or not with the presence of temporal adverbs expressing change and the first person); the use of the Imperfect tense alone, we add, is an index of the typological context. De Jonge's contribution is fundamental because it links the contextual expansion of *estar* to a combination of pragmatic meaning (the expression of a contrastive situation connected to the speaker's perspective) and syntactic markers (that can be considered as indexes of different degrees of innovation with respect to the use of *estar* in age predications).<sup>5</sup>

The subjective meaning of the innovative *estar* sentences had also been acknowledged in Gutiérrez (1992), who claimed that the innovative construction expresses the speaker's subjective evaluation of the attribution of the property; i.e. the predication is presented from a subjective point of view. This intuition has been reformulated in different ways in more recent analyses that argue that the basic meaning of the copula *estar* is the same in standard and innovative constructions. For example,

---

<sup>5</sup> It is important to remark that De Jonge claims that the synchronic distribution of *estar* must be understood as a reflex of a diachronic process of extension from the standard use (as also claimed in Malaver 2009, Piñango and Fuchs 2023, a.o.). The diachronic perspective will be set aside in this paper.

Escandell-Vidal and Leonetti (2016) and Escandell-Vidal (2018, 2023) proposed that <*estar* + age adjective> sentences have an experiential meaning that arises from a semantic mismatch between the copula *estar* and the adjective. In a sentence like *Juan está joven* ('Juan is young'), *estar* requires a context-dependent situational predicate and *joven* is a property adjective (individual-level predicate), which provokes a semantic clash. The pragmatic component resolves the incompatibility by introducing a situation that satisfies the semantic requirements of *estar*: the situation of acquisition of evidence by the speaker, thus giving rise to the experiential, hence subjective, interpretation of the predication (in all Spanish varieties, so that dialectal differences are not well explained).

Similarly, Sánchez-Alonso, Deo and Piñango (2019) claim that *estar* requires the predication to be linked to a situation (defined with respect to an evaluation index *i*), introducing the presupposition of an alternative situation (defined with respect to an evaluation index *i'*) in which the predication does not hold. An evaluation index *i* is composed of the parameters  $i\langle t,w,l,c \rangle$ —*t*: time interval, *w*: world, *l*: location, *c*: function that assigns the positive degree of the gradable predicate—. The innovative use of *estar* (as in *La Capilla Sixtina está hermosa* 'The Sixtine Chapel is beautiful', as an expression of aesthetic judgment upon seeing it for the first time) is related to the modal parameter *w*, which defines an alternative world in which the ascription of the property to the subject by the speaker could be judged differently. What characterizes innovative varieties is the possibility of accommodating that presupposition without syntactic cues that trigger it.

Moreno-Quibén (2022) and Gumiel-Molina, Moreno-Quibén and Pérez-Jiménez (2023, 2024) also defend a uniform semantics for *estar* in standard and innovative varieties and have as a core empirical fact the subjective/experiential meaning component of the innovative construction. These authors link this meaning component to lexical-syntactic properties of adjectives, given the differences found in Figure 1 above. Specifically, for the macro-class of evaluative adjectives (6), they argue that certain subclasses of evaluative adjectives undergo a geographically conditioned process of argument augmentation in innovative varieties (dative experiencer augmentation). This argument provides the stage necessary to ground the predication with *estar* and gives rise to the meaning of potentially changing situation relevant to the speaker, which is coreferential with the dative. This proposal explains why aesthetic adjectives (like *hermosa* 'beautiful', in the example above) are possible in the innovative construction in some Spanish varieties, where the predication does not imply a change in the subject with respect to the property: aesthetic adjectives, as a natural class of evaluative adjectives,

have an enriched argument structure in innovative Spanish varieties. It also nicely explains the different incidence rate of appearance of other natural classes of adjectives in the innovative construction (for example, dimensional adjectives, like *alto* ‘tall’). However, with respect to innovative *estar* sentences with age adjectives, although subjectivity/evidentiality could be a meaning component of the examples illustrating neutral contexts, it does not seem to be clearly the case in the typological contexts, (2), (3), where scalar variation and faultless disagreement are not possible, as will be shown below; note that (2) is a purely objective ascription of an individual into a class defined by their age, in which there is also no functional standard that could be a source of subjectivity. This could indicate that the innovative appearance of age adjectives with the copula *estar* has a different source.

In the following section, we will present our database and the results regarding the current distribution of *ser* and *estar* with age adjectives across Spanish dialects and different kinds of syntactic-semantic contexts.

### 3. DATABASE. RESULTS

To describe the behavior of *ser* and *estar* with adjectives of age across Spanish varieties, we built a database based on the PRESEEA corpus (Proyecto para el Estudio Sociolingüístico del Español de España y de América, <https://preseea.uah.es>). For each city represented in the corpus, there are 18 semi-directed oral interviews, with text-audio alignment (with a total of total 5,259,254 words). The corpus is designed following common methodological criteria—sample size, sociological stratification, and transcription protocol—which guarantees the comparability of data across dialectal varieties. The oral texts were produced by male and female speakers of three age groups (20–34; 35–54; 55+) and three levels of education (low, medium, high) (Moreno Fernández 2021). The cities surveyed are representative of the different dialectal areas of Spanish, as defined in Moreno Fernández (2019, a.o), RAE-ASALE (2009, 2025), Malaver (2022) and Orozco (2022):<sup>6</sup>

---

<sup>6</sup> A clarification is in order regarding the areas of Central America and Mexico; these are considered a single area in Moreno Fernández (2009), and RAE-ASALE (2009, 2025), among others, but are considered two different areas in the later work of Moreno Fernández (2019). In this work we considered them as a single area for methodological reasons: the Central American area is represented in the PRESEEA corpus only by the Guatemala corpus (with 122,478 words, compared to, for example, the Andean corpus, which

- **Mexican and Central American Area:** Mexico (Mexicali MXLI, Mexico City MEXI, Guadalajara GUAD, Monterrey MONR, Puebla PUEB); Guatemala (Ciudad de Guatemala GUAT): 108 interviews, 1,280,631 words (1,158,153 words in Mexico and 122,478 in Central America). Data collected between 2001 and 2018.
- **Caribbean Area (Continental and Antillean):** Colombia (Barranquilla BARR), Venezuela (Caracas CARA), Cuba (La Habana LHAB): 54 interviews, 479,038 words (2001–2011).
- **Andean Area:** Bolivia (La Paz LPAZ), Colombia (Bogotá BOGO, Cali CALI, Medellín MEDE, Pereira PERE), Perú (Lima LIMA), Venezuela (Mérida MEVE): 126 interviews, 969,863 words. (2004–2022)
- **Chilean Area:** Chile (Santiago de Chile SCHI): 18 interviews, 208,938 words. (2007– 2009).
- **Rioplatense Area:** Argentina (Buenos Aires BAIR), Uruguay (Montevideo MONV): 36 interviews, 339,590 words. (2007–2012).
- **European Area:** Spain (Madrid MADR, Alcalá de Henares ALCA, Gijón XIXO, Granada GRAN, Cádiz CADI, Las Palmas de Gran Canaria LASP, Malaga MALA, Santander SANT, Santiago de Compostela SCOM, Seville SEVI, and Valencia VALE): 198 interviews, 1,981,194 words. (1988 – 2023)

The search strings used to extract examples of [*copula* + adjective of age] are the following (the same strings were used with lemma='ser'):

- Search 1: any form of the verb *estar* / *ser* + an adjective  
 ([[lemma='estar' & pos='V.+'%c]] [[pos='AQ.+'%c]])
- Search 2: any form of the verb *estar* / *ser* + an adverb + an adjective  
 ([[lemma='estar' & pos='V.+'%c]] [[pos='R.+'%c]] [[pos='AQ.+'%c]])
- Search 3: sequences *qué* + adjective + any form of the verb *estar* / *ser*  
 [(word='qué'%c)] [[pos='AQ.+'%c]] [[lemma='estar'%c]]

After eliminating repetitions and non-pertinent examples (for example, sentences containing adjectives like *pequeño* 'small' expressing size), a total of 918 valid cases were obtained: 472 (51.42%) examples of *ser*+age adjective; 446 examples of *estar*+age adjective (48.58%). The contexts were expanded to include complete paragraphs so that the syntactic parameters described below could be annotated and the contexts could be appropriately interpreted. The age adjectives documented in the examples in the database are *adolescente* ('adolescent'), *adulto* ('adult'), *chavo* ('young'), *chico*, *chiquitito*,

---

has 969,863 words). This imbalance could pose a problem for the statistical analysis so that Mexico and Central America were considered together.

*chiquito* ('young'), *grande* ('old'), *joven* ('young'), *mayor* ('old'), *niño*, *pequeño* ('little'), *sardino* ('young') and *viejo* ('old').

Table 2 shows the distribution of *ser* and *estar* + age adjective in the different varieties of Spanish.<sup>7</sup> As can be seen, although the total number of examples with each copula is fairly similar (472 *ser* – 446 *estar*), their distribution across the different dialectal areas is uneven (the geographical distribution of the different age adjectives will not be explored in this paper, see Malaver 2009, Gumiel-Molina, Moreno-Quibén and Pérez-Jiménez 2024). The copula *ser* predominates in areas that have been considered in the literature as conservative with regard to the presence of the innovative construction with *estar* (European Spanish: 92% *ser* - 7.25% *estar*; Rioplatense area: 88.37% *ser* - 11.63% *estar*). The copula *ser* also appears predominantly in Chilean Spanish (78% *ser* vs. 22% *estar*). On the contrary, the copula *estar* predominates with age adjectives, as expected, in the so-called innovative Spanish varieties. Thus, in Mexico and Central America, only 25.45% of the examples contain *ser* compared to the 74.55% of *estar* examples; the Andean and Caribbean areas show similar incidence rates of each copula: 38.99% / 38,83% (*ser*) vs. 61.01% / 61,16% (*estar*).<sup>8</sup>

Variety	Ser	%	Estar	%	TOTAL
Mexican & Central American	71	25.45	208	74.55	279
Caribbean	47	38.83	74	61.16	121
Andean	85	38.99	133	61.01	218

<sup>7</sup> While traditional frequentist tests of independence (such as the chi-square test) or measures of association (such as Cramer's V) are commonly used for categorical count data, they evaluate raw counts and can be biased by highly unequal sample sizes across dialectal areas (e.g., 279 observations in Mexico and Central America versus 43 in the Rioplatense area). Therefore, the present article deliberately employs a Bayesian Poisson regression framework to account for the specific nature of the data, as this approach is particularly well-suited for modeling count frequencies while simultaneously considering varying sample sizes. By integrating an offset term (*logNRegion*), the model normalizes the total number of observations per region, thereby modeling proportions rather than raw counts. We ensure that areas with smaller data samples do not exert an undue influence on the overall parameter estimates. See the illuminating discussion in chapter 1 of McElreath (2020). We would like to thank one of our anonymous reviewers for compelling us to clarify this point.

<sup>8</sup> Gumiel-Molina, Moreno-Quibén and Pérez-Jiménez (2024: 11) showed that 98.6% of examples with age adjectives in American varieties illustrate innovative use. Examples illustrating standard use are concentrated in the Rioplatense area.

Chilean	39	78	11	22	50
Rioplatense	38	88.37	5	11.63	43
European	192	92.75	15	7.25	207
Total	472	51.42	446	48.58	918

Table 2. *Ser* and *Estar* + age adjective across dialectal areas

As noted in the previous section, the innovative *estar* contexts can be grammatically identified. Neutral contexts generally contain frame temporal expressions (e.g. *cuando*), with the concurrence, in many cases, of explicit syntactic markers typical of the standard chronological use, such as modifiers expressing temporal phase (*ya, todavía*), or first-person subjects, referring to the speaker's perspective. Typological contexts are consistent with the use of certain verb tenses such as the Imperfect. Accordingly, the examples in the database were annotated with respect to the following parameters:

- +presence of a temporal adverb expressing change, namely *ahora, ya, todavía*, etc.
- +person: 1, 2, 3 (sg and pl)
- +type of sentence: subordinate sentence expressing a time frame (*cuando*) vs. other syntactic contexts.
- +verbal tense: Present, Imperfect, Gerund, others (Future, Present perfect, etc.).

Let us begin with the analysis of the presence of *ser* and *estar* across Spanish varieties in first-person predications (3.1) and in sentences with temporal expressions expressing change (3.2). The distribution of *ser* and *estar* by type of sentence (*cuando* vs. others) will be analyzed in 3.3. Finally, the relation between the appearance of *ser/estar* and verbal tense will be dealt with in 3.4. We will take the distribution of these grammatical features in European Spanish as the baseline for comparing statistically the distributions (see section 4).

### 3.1. <Ser/Estar + Adjectives of Age> and the Use of the First Person

The distribution of *ser* and *estar* in first-person predications across dialectal varieties is shown in Table 3.

Variety	SER	%	ESTAR	%	TOTAL
---------	-----	---	-------	---	-------

Mexican & Central American	20	20.83	76	79.17	96
Caribbean	4	12.12	29	87.88	33
Andean	33	39.76	50	60.24	83
Chilean	6	46.15	7	53.85	13
Rioplatense	15	78.95	4	21.05	19
European	62	92.54	5	7.46	67
Total	140		171		311

Table 3. *Ser* and *Estar* in First-Person Predications

As can be seen in Table 3, the incidence rate of *ser* and *estar* in first-person contexts in European Spanish is the same as in the overall data set for this area (Table 2) (92.54% first-person predications with *ser* – 7.46% with *estar*); furthermore, the examples of first-person predications with *estar* conform to the standard use of this copula, as shown in (14). A similar trend can be observed in the Rioplatense area (78.95% first-person predications with *ser* - 21.05% with *estar*). In the Mexican-Central American and Andean varieties, the incidence rate of *ser* and *estar* in first-person predications is also consistent with their rate of appearance in the entire database: in Mexico and Central America, 20.83% of first-person predications contain the copula *ser* compared to 79.17% that contain *estar*; in the Andean area, 39.76% of the examples contain *ser* compared to 60.24% with *estar*. In the Chilean area, where the total number of examples with *ser* is higher in the database (78% *ser* – 22% *estar*), however, 46.15% of first-person predications contain the copula *ser* and 53.85% contain *estar*. The data for the Caribbean area is especially relevant, where 12.12% of the first-person examples contain *ser* compared to 87.88% that contain innovative *estar*, (15) (compared to the general distribution of the copulas in the database 38.83 % *ser* – 61,16 % *estar*).

- (14) ya estoy bostezando porque me da un sueño <risas title="A1, E"> no aguanto yo ya estoy vieja <risas title="E"> ¿y qué es discotecas así con musica latina especializada o así? Sí [XIXO\_M11\_049]

“I'm already yawning because I'm so sleepy <laughs title="A1, E"> I can't take it anymore. I'm<sub>ESTAR</sub> already old <laughter title="E"> And what are clubs like that with specialized Latin music or something? Yes”

- (15) él compró un tocadisco (...) ¿tú sabes cuál fue la primera canción que él estrenó ahí en / en / en ese tocadisco? / este / Juancito Trucepei // ¡lo compró en disco de setenta y ocho revoluciones! / **cuando yo estaba joven** / cuando yo estaba pe <palabra\_cortada/> primero te estoy hablando de **cuando estaba niña** / cuando estaba niña / existían solamente treinta y ocho revoluciones / después apareció las cuarenta y cinco revoluciones [CARA\_M33\_103]
- “He bought a record player / called a picó / (...) Do you know what was the first song he playd on / on / on that record player? / This one / Juancito Trucepei // He bought it on a 78 rpm record! / when I was<sub>ESTAR</sub> young / when I was<sub>ESTAR</sub> pe <word\_cut off/> first I'm talking about when I was<sub>ESTAR</sub> a girl / when I was a girl / there were only thirty-eight revolutions / then the forty-five revolutions appeared”.

The data therefore show that, except in the Andean region and Chile, the percentages of predications with *ser* and *estar* in the first person are the same as those found in the entire corpus for all Spanish varieties (Table 2). This is expected if this grammatical feature is linked both to the chronological contexts in which *estar* appears and to neutral (innovative) contexts. The statistical analysis will show that this factor seems not to be relevant to explain the distribution of *ser* and *estar* across Spanish varieties: Malaver (2009) also claimed that there is no significant association of the copula *estar* with the first person to any particular variety.

### 3.2. <Ser/Estar + Adjectives of Age> with Temporal Expressions

Table 4 shows the distribution of *ser* and *estar* in sentences with temporal expressions denoting change (namely *ya*, *todavía*) across dialectal varieties.

Variety	Ser	%	Estar	%	TOTAL
Mexican & Central American	16	14.68	93	85.32	109
Caribbean	12	31.58	26	68.42	38

Andean	16	38.10	26	61.90	42
Chilean	5	50	5	50	10
Rioplatense	7	70	3	30	10
European	37	82.22	8	17.78	45
Total	74		158		232

Table 4. Combination of <Ser/Estar + Adjective of Age + Temporal Expression>

In the Rioplatense area, the distribution of *ser* and *estar* in sentences containing *ya* (70% *ser* – 30% *estar*) parallels the distribution found in first-person predications; European Spanish shows a similar tendency (82% *ser* – 18% *estar*). The results for the Andean (38% *ser* – 61.9% *estar*) and Chilean areas (50% *ser* – 50% *estar*) mimic those described in Table 3 for first-person predications. The results, however, are not parallel to those obtained in Table 3 for the other areas. In Mexico and Central America, 14.68% of *ya*-predications contain the copula *ser* compared to 79.17% that contain *estar*. In the Caribbean area 31.58% of the examples with the temporal marker *ya* contain *ser* compared to 68.42% that contain innovative *estar*. As we will show, however, these differences are not statistically significant, except for the Andean area. This is also expected if this grammatical feature is linked both to the chronological contexts in which *estar* appears and to neutral (innovative) contexts.

### 3.3. <Ser/Estar + Adjectives of Age> in Cuando Clauses

The distribution of the copulas *ser* and *estar* in sentences headed by *cuando* is shown in Table 5.

Variety	Ser	%	Estar	%	TOTAL
Mexican & Central American	21	40,38	31	59,62	52
Caribbean	26	36,36	59	63,64	85
Andean	8	30,59	14	69,41	22
Chilean	6	60	4	40	10
Rioplatense	7	100	0	0	7

European	29	100	0	0	29
Total	97		108		205

Table 5. <Ser/Estar + adjectives of age> in sentences with *cuando* ('when')

The relative adverb *cuando*, as stated above, prototypically identifies the contexts that De Jonge calls neutral (it also appears in innovative *estar* sentences with typological meaning). The data show that in the Rioplatense and European areas, the presence of *estar* in *cuando* sentences is not documented. On the contrary, *estar* sentences with *cuando* are documented in the rest of the American varieties, with difference incidence rates. The results for the Andean (30.59% *ser* – 69.41% *estar*) and Caribbean areas (36.36% *ser* – 63.64% *estar*) are in line with those described in table 4 for *ya* predications. In Chile, 60% of the examples with *cuando* contain *ser*; 40% contain *estar*. In Mexico and Central America, 40.38% of *cuando* sentences contain the copula *ser* compared to 59.62% that contain *estar*. As we will see in section 4, there is a significant correlation between the presence of *cuando* and geolectal areas, which can thus be grouped based on this parameter.

### 3.4. Verbal tense

Finally, Tables 6 and 7 show the distribution of verb tenses in sentences with *ser* and *estar*. Anticipating the statistical analysis, the Imperfect tense is the only significant tense for the distribution of copulative verbs across Spanish varieties. It groups together the European and Rioplatense varieties (in which no sentences with *estar* in the Imperfect tense are documented) as opposed to the rest of the American varieties. The Mexico-Central American and Andean areas stand out in the appearance of sentences with *estar* in the Imperfect tense, followed by the rest of the American dialectal areas. Other tenses do not establish differences with respect to the use of *ser* and *estar* among the different varieties of Spanish.

Variety	PRES	%	IMP.	%	GERUND	%	Others	%
Mex. & Central Am.	51	76,12%	16	22,86%	0	0,00%	3	4,29%
Caribbean	49	60,49%	30	35,29%	2	2,35%	4	4,71%
Andean	33	75,00%	9	20,00%	2	4,44%	1	2,22%
Chilean	35	94,59%	2	5,13%	0	0,00%	2	5,13%
Rioplatense	22	66,67%	11	32,35%	0	0,00%	1	2,94%

European	133	70,37%	33	17,46%	13	6,88%	10	5,29%
Total	323	68,43%	101	21,40%	17	3,60%	21	4,45%

Table 6. Verbal tense in <ser + adjective of age> sentences

Variety	PRES.	%	IMP.	%	GERUND	%	Others	%
Mex. & Central Am.	100	48,08%	106	50,96%	2	0,96%	0	0,00%
Caribbean	43	32,58%	86	65,15%	3	2,27%	0	0,76%
Andes	36	48,65%	37	50,00%	1	1,35%	0	0,00%
Chile	4	40,00%	6	54,55%	0	0,00%	1	9,09%
Rioplatense	5	100%	0	0,00%	0	0,00%	0	0,00%
Spain	13	86,67%	1	6,67%	1	6,67%	0	0,00%
Total	201	45,07%	235	52,69%	6	1,35%	2	0,45%

Table 7. Verbal tense in <estar + adjective of age> sentences

Malaver (2009) finds that the Imperfect tense is preferentially combined with *estar* in all the American areas analyzed, in contrast to what happens in European Spanish. In some areas, such as Mexico, the association is statistically significant, which means that the Imperfect is not independent of *estar* in the age predications (8.5.1.3). Malaver links the high incidence of the Imperfect tense in general in the corpus to the fact that speakers use it in their narratives to answer typical questions, such as “Do you think the city has changed a lot?” or “What was your childhood like?” However, its geographical distribution with copulas is crucially different.

#### 4. A BAYESIAN APPROACH TO MODELING THE DISTRIBUTION OF <ESTAR+AGE ADJECTIVES> BASED ON LINGUISTIC FEATURES

To analyze the significance of the linguistic factors introduced in the preceding section for the distribution of the verbs *ser* and *estar* across different Spanish-speaking regions, a Bayesian Poisson regression framework was employed (see McElreath 2020, Gellman & Hill 2021). This approach is well-suited for modeling count data (frequencies of linguistic features) while accounting for varying sample sizes across regions (recall footnote 3). The justification for adopting this framework stems from its inherent suitability for modeling the type of data collected: count data, specifically the observed frequencies of occurrences of *ser* and *estar* in various linguistic contexts (as justified in Vasishth & Nicenboim 2016 and Nicenboim & Vasishth 2016). By integrating regional

sample size or subcorpus token count within the bayesian structure, the model ensures that regions with smaller data samples do not exert an undue influence on the overall parameter estimates, thus lending greater stability to the comparative analysis.

The model was specified as follows as:<sup>9</sup>

$$\log(\lambda_i) = \alpha + \beta_{\text{Region}} + \beta_{\text{Feature}} + \beta_{\text{Region} \times \text{Feature}} + \log(N_{\text{Region}})$$

The model specification was formulated to capture the variation as a function of key explanatory variables (such as grammatical features and specific geographic regions) while accounting for the count nature of the dependent variable (the frequency of the verb's use). In this model,

- $\lambda_i$  represents the expected count of a specific feature.
- $\alpha$  represents the intercept of the model. That is, it represents the baseline (log-) frequency of the reference feature in the reference region (Spain).
- The **interaction term** ( $\beta_{\text{Region} \times \text{Feature}}$ ) is the parameter of interest, quantifying how the usage of a specific feature in a given region deviates from the baseline.
- **Spain (Peninsular Spanish)** was selected as the reference baseline ( $\beta_{\text{Region=Spain}} = 0$ ).
- An **offset term** ( $\log(N_{\text{Region}})$ ) was included to normalize for the total number of observations per region, effectively modeling rates/proportions rather than raw counts.

Models were fitted separately for each of the key grammatical categories under investigation: **Person** (*Persona*), capturing distinctions such as first, second, and third person; **Temporal Expression** (*Expresión Temporal*), presence vs. absence of temporal adverbials expressing change such as *ya* (“already”), *todavía* (“then”), etc. **Sentence Type** (*Tipo de Oración*), which classified structures with *cuando* and *sin\_cuando*; and **Tense** (*Tiempo*), encompassing Present, Imperfect, Gerund and others (future forms; compound perfect tenses).

The statistical significance of the findings, particularly regarding regional variations or deviations from a baseline, was determined by inspecting the 95% Credible Intervals (CI) of the posterior distributions resulting from the Bayesian modeling

---

<sup>9</sup> In order to test the model, we implement it in R Statistical Software (4.5.2 R Core Team 2025) using the package `rstanarm` (Goodrich *et al.* 2025, see also Sorensen, Hohenstein, & Vasishth 2016). Graphics were plotted in R with the package `ggplot` (Wickham 2016, Wickham et al. 2021).

approach. A regional deviation or interaction term was deemed statistically significant, indicating a genuine effect beyond random chance, only if the corresponding 95% Credible Interval completely excluded zero. This criterion ensured a robust identification of significant regional differences in the use of these specific grammatical features.

We provide here an implementation of the model to a specific linguistic feature: the Imperfect tense with *estar* (*estaba*) in the Andean region. The model estimates the log-expected count ( $\log(\lambda_i)$ ) using the following coefficients derived from the data. The intercept, ( $\alpha$ ), the baseline log-rate of the reference feature (Present Tense) in the reference region (Spain), equals -0.1431. The parameter  $\beta_{\text{Region}}$  (Andean region) represents the “penalty” for being in the Andean region wrt. to the reference region (Spain) for the reference feature. It accounts for differences in sample size or baseline usage of “Presente” (‘Present’) and equals -0.9860. The parameter  $\beta_{\text{Feature}}$  (Imperfecto ‘Imperfect’) represents the drop in frequency when moving from “Presente” to “Imperfecto” in the baseline region (Spain). Since Imperfecto with *estar* and adjective of age is rare in Spain, this number is strongly negative, -2.5649. The interaction term,  $\beta_{\text{Region} \times \text{Feature}}$  (Andean x Imperfecto), represents the “boost” for areal innovation, 3.2581. The linguistic interpretation of this magnitude is that using *estar* in the Andean area drastically increases the probability of using the Imperfect tense (*estaba*) compared to what we would expect from the Spanish baseline. Finally,  $\log(N_{\text{Region}})$  is 4.8903, where  $N = 133$  is the total number of *estar* tokens in the Andean count. Hence,

$$\begin{aligned} \log(\lambda_i) &= \alpha + \beta_{\text{Andean}} + \beta_{\text{Imperfecto}} + \beta_{\text{Andean} \times \text{Imperfecto}} + \log(N_{\text{Andean}}) \\ \log(\lambda_i) &= -0.1431 + (-0.9860) + (-2.5649) + 3.2581 + 4.8903 \\ \log(\lambda_i) &= 4.4544 \\ \lambda &= 86 \end{aligned}$$

This example demonstrates that the **Interaction Term** (+3.2581) is the key descriptive finding behind the result. Without it, the model would predict a very low count for Imperfecto (inheriting the negative -2.56 from Spain). The large positive interaction corrects this, accurately capturing that the Andean variety has "promoted" the Imperfect tense with *estar* to a major variant, characterizing the linguistic area. For the feature **Person** (*Persona*), we fitted the model as show in Table 8.

Term	Estimate	Std. Error	Z value	Pr(< z )	Risk Ratio	Sig.
Intercept (Baseline: Spain, 3 <sup>a</sup> )	-0,5108	0,3333	-1,5325	0,1254	0,6	
Feature: 1 <sup>a</sup>	-0,5878	0,5578	-1,0538	0,292	0,5556	
Feature: 2 <sup>a</sup>	-23,7868	29577,8	-0,0008	0,9994	0	
Region: Andean	0,0149	0,3514	0,0425	0,9661	1,015	
Region: Caribbean	-0,1297	0,3698	-0,3507	0,7258	0,8784	
Region: Chilean	-0,5008	0,6009	-0,8333	0,4047	0,6061	
Region: Mx&Cent	-0,496	0,3525	-1,407	0,1594	0,609	
Region: Rioplaten	-1,0986	1,0541	-1,0422	0,2973	0,3333	
Region: Andean : Feature: 1 <sup>a</sup>	0,1054	0,5861	0,1798	0,8573	1,1111	
Region: Caribbean : Feature: 1 <sup>a</sup>	0,2915	0,6093	0,4785	0,6323	1,3385	
Region: Chilean : Feature: 1 <sup>a</sup>	1,1474	0,839	1,3675	0,1715	3,15	
Region: Mx&Cent : Feature: 1 <sup>a</sup>	0,5878	0,5809	1,0119	0,3116	1,8	
Region: Rioplaten : Feature: 1 <sup>a</sup>	1,9741	1,2494	1,58	0,1141	7,2	

Table 8. (Significance codes: \*\*\* < 0.001, \*\* < 0.01, \* < 0.05, . < 0.1)

Unlike the Tense category, where we will see massive shifts (Risk Ratios > 20), the interaction terms for Person are largely **non-significant** ( $p > 0.05$ ). This statistically confirms the finding mentioned in the previous section: the variable **Person** does not strongly influence the regional distribution of *estar* compared to the baseline of Spain. The usage of 1st vs 3rd person remains relatively stable across regions for this verb.

With regard to the category **Temporal Expression** (*Expresión Temporal*), our model fitted to the following values in Table 9.

Term	Estimate	Std. Error	Z value	Pr(< z )	Risk Ratio	Sig.
Intercept (Baseline: Spain, Sin exp. tpal)	-0.7621	0.378	-2.0164	0.0438	0.4667	*
Feature: Con exp. tpal	0.1335	0.5175	0.258	0.7964	1.1429	
Region: Andean	0.5352	0.3902	1.3715	0.1702	1.7078	
Region: Caribbean	0.3293	0.4046	0.8139	0.4157	1.39	
Region: Chilean	0.156	0.5563	0.2804	0.7792	1.1688	
Region: Mx&Cent	0.1695	0.3893	0.4355	0.6632	1.1848	
Region: Rioplatense	-0.1542	0.8018	-0.1923	0.8475	0.8571	
Region: Andean: Feature: Con exp. Tpal	-1.5389	0.5619	-2.7386	0.0062	0.2146	**
Region: Caribbean: Feature: Con exp. Tpal	-0.7466	0.572	-1.3054	0.1918	0.474	

Region: Chilean: Feature: Con exp. Tpal	-0.3159	0.7966	-0.3965	0.6917	0.7292	
Region: Mx&Cent: Feature: Con exp. Tpal	-0.3459	0.536	-0.6453	0.5188	0.7076	
Region: Rioplaten: Feature: Con exp. Tpal	0.2719	1.0494	0.2591	0.7955	1.3125	

Table 9. (Significance codes: \*\*\* < 0.001, \*\* < 0.01, \* < 0.05, . < 0.1)

The **Andean** region shows a significant negative interaction with the temporal expression feature ( $p = 0.0062$ ). While the baseline (Spain) uses *estar* with temporal expressions at a certain rate (Risk Ratio  $\sim 1.14$  vs. 0.446 without), the Andean region uses temporal expressions significantly **less frequently** relative to that baseline behavior (Risk Ratio  $\sim 0.21$  for the interaction). Andean speakers are much more likely to use *estar* **without** temporal expressions compared to the Spanish standard.

The feature **Sentence Type** (Tipo de Oración) which makes reference to sentences without *cuando* and with *cuando* was fitted with the following values as shown in Table 10.

Term	Estimate	Std. Error	Z value	Pr(< z )	Risk Ratio	Sig.
Intercept (Baseline: Spain, Sin CUANDO)	0	0.2582	0	1	1	
Feature: Cuando	-23.7858	22899.4	-0.001	0.9992	0	
Region: Andean	-0.5863	0.2832	-2.0705	0.0384	0.5564	*
Region: Caribbean	-0.2097	0.2887	-0.7265	0.4675	0.8108	
Region: Chilean	-0.452	0.4577	-0.9874	0.3234	0.6364	
Region: Mx&Cent	-0.1614	0.2689	-0.6001	0.5484	0.851	
Region: Rioplatense	0	0.5164	0	1	1	
Region: Andean : Feature: Cuando	23.5593	22899.4	0.001	0.9992	1700000000	
Region: Caribbean : Feature: Cuando	22.3305	22899.4	0.001	0.9992	4900000000	
Region: Chilean : Feature: Cuando	23.2262	22899.4	0.001	0.9992	12000000000	
Region: Mx&Cent : Feature: Cuando	22.0436	22899.4	0.001	0.9992	3700000000	
Region: Rioplaten : Feature: Cuando	1.0986	32384.7	0	1	3	

Table 10. (Significance codes: \*\*\* < 0.001, \*\* < 0.01, \* < 0.05, . < 0.1)

The statistical model encounters a mathematical problem here because the feature *cuando* is completely absent in the Spain count (0 occurrences). This causes the "Estimate" for Feature: Cuando to go to negative infinity (represented here as  $\sim -23.78$ ). While the p-values appear non-significant (approx. 1.0) due to the inflated standard errors caused by the zero count, the "Risk Ratio" column offer a different perspective: use of *estar* in *cuando* clauses with adjectives of age is a prevalent feature in the linguistic landscape of American varieties of Spanish (except for the Rioplatense area). It exists there (e.g., Andean Risk Ratio is astronomically high relative to zero), whereas it simply does not exist in the Peninsular baseline data.

The verb *ser* maintains a **relatively stable usage pattern** throughout the surveyed regions as can be seen in Figure 2. Regional deviations in its frequency and contextual application generally remain contained within a narrow band—specifically, within  $\pm 10$  percentage points of the established European Spanish baseline. This uniformity suggests a deep, underlying structural consensus regarding the fundamental role of *ser* in the Spanish verbal system.

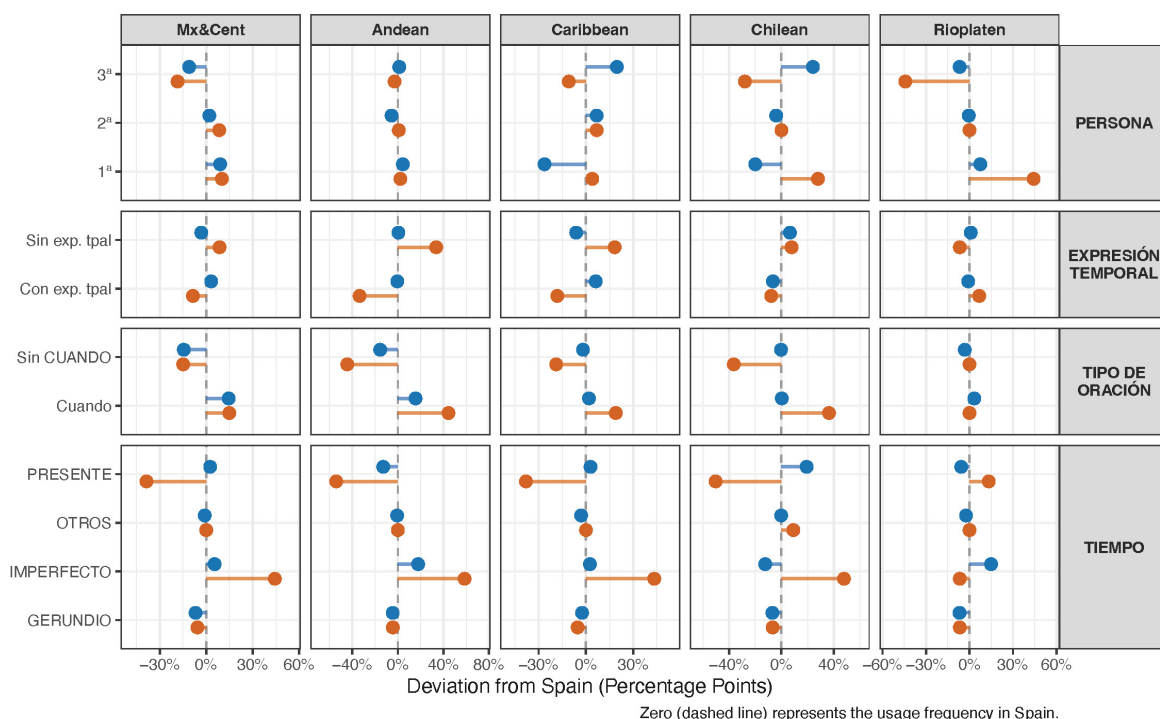


Figure 2. Comparative deviation: *ser* (upper dot of each pair) vs. *estar* (lower dot of each pair). The distance between the upper (*ser*) and the lower dot (*estar*) highlights where the two verbs behave differently. The dashed vertical line indicates the Spanish baseline.

In contrast, the verb *estar* displays significant regional variation, most saliently when the data is broken down by the Tense (*Tiempo*) category, Table 11. The most compelling and statistically significant finding of this analysis is the substantial increase in the usage of the Imperfect tense (*estaba*) for *estar* and adjectives of age within Latin American varieties of Spanish. This phenomenon is particularly pronounced in two major linguistic zones: Mexico & Central America and the Andean region. Note in Figure 2 the extreme rightward shift of the red dots for "Imperfecto" in the Mexican and Andean panels, indicating a massive overuse relative to Spain.

Term	Estimate	Std. Error	Z value	Pr(< z )	Risk Ratio	Sig.
Intercept (Baseline: Spain - Presente)	-0,1431	0,2774	-0,516	0,6059	0,8667	
Feature: GERUNDIO	-2,5649	1,0377	-2,4716	0,0134	0,0769	*
Feature: IMPERFECTO	-2,5649	1,0377	-2,4716	0,0134	0,0769	*
Region: Andean	-0,986	0,3165	-3,1154	0,0018	0,373	**
Region: Caribbean	-0,5774	0,3236	-1,7846	0,0743	0,5613	.
Region: Chilean	-0,8685	0,5718	-1,519	0,1288	0,4196	
Region: Mx&Cent	-0,5893	0,2948	-1,9987	0,0456	0,5547	*
Region: Andean : Feature: IMPERFECTO	3,2581	1,0544	3,0899	0,002	26	**
Region: Caribbean : Feature: IMPERFECTO	2,5923	1,0638	2,4368	0,0148	13,3611	*
Region: Chilean : Feature: IMPERFECTO	2,9704	1,2221	2,4305	0,0151	19,5	*
Region: Mx&Cent : Feature: IMPERFECTO	2,6232	1,0471	2,5053	0,0122	13,78	*

Table 11. (Significance codes: \*\*\* < 0.001, \*\* < 0.01, \* < 0.05, . < 0.1)

The contrast in tense preference with adjectives of age is clear when comparing the core regional clusters. In European Spanish, usage of *estar* with adjectives of age is overwhelmingly dominated by the Present tense (*estoy/está*). The Imperfect forms contribute only a negligible proportion to the overall frequency profile of the verb. In the Latin American varieties, except for the Rioplatense area, which aligns with Spain, the Imperfect tense (*estaba*) with adjectives of age emerges as prominent. As graphically detailed in Figure 3, the frequency of Imperfect forms within the data for Mexico and the Andes is demonstrably comparable to, or even surpasses, that of the Present tense. If we

consider the Imperfect to be a sign of *estar* appearing in innovative contexts (neutral contexts, in combination with *cuando*, and innovative typological contexts), then this distribution clearly draws a line between non-innovative and innovative areas regarding the use of *estar*. Moreover, within the innovative group, it distinguishes between areas with more advanced innovation and those with more moderate innovation.

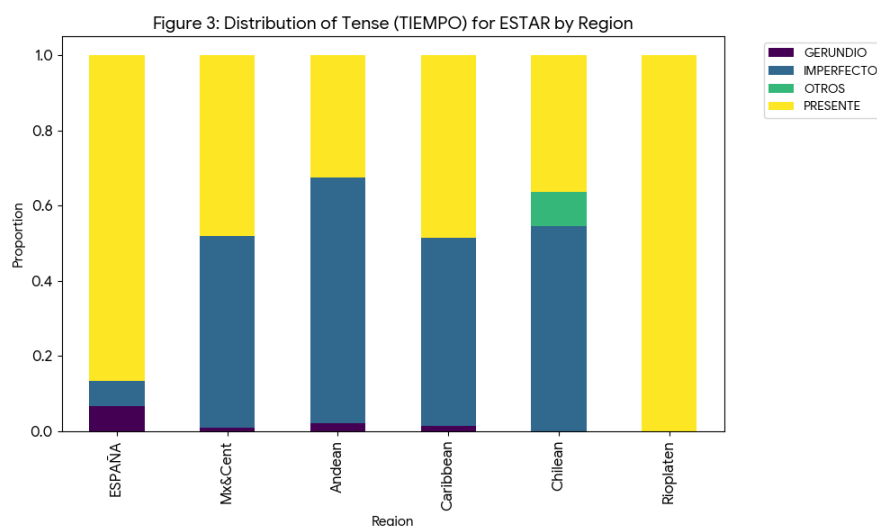


Figure 3. Imperfect tense vs. Present tense with *estar* and adjective of age

The robustness of this dialectal feature is confirmed by quantitative analysis. The interaction effect for **Region x Feature:Imperfecto** yielded a highly positive and statistically significant result. This statistical confirmation solidifies the Imperfect Shift as a robust dialectal feature —a systematic, language-internal divergence— rather than merely a product of random linguistic fluctuation or sampling. The influence of the Imperfect tense extends beyond mere frequency; it signals a semantic and syntactic shift in how *estar* is functioning along various axis. In standard Spanish, *estar* with adjectives of age is typically used in chronological contexts (referring to a specific moment). However, the use of the Imperfect tense (*estaba*) in innovative varieties often signals a typological context in De Jonge’s term. In these contexts, the speaker is not just describing a temporary stage but is ascribing an individual to a class (recall example (2)). The Imperfect tense serves as a syntactic clue that *estar* is taking over the classifying function traditionally held by *ser*, indicating a more advanced stage of grammaticalization. This use of the Imperfect allows for objective descriptions of age in the past without necessarily invoking the subjective or experiential meaning often associated with innovative *estar* in the Present tense.

## 5. CONCLUSIONS

This study has employed a Bayesian regression framework to provide an updated descriptive and theoretical analysis of the distribution of *ser* and *estar* with age adjectives across Spanish dialects. The statistical analysis of the Preseca corpus yields a clear geolectal hierarchy regarding the innovative use of *estar* with this class of adjectives. The most innovative varieties are found in Mexico, Central America, and the Andean region, where the use of *estar* is fully established and correlates strongly with specific syntactic factors which function as markers of contexts of use more (neutral contexts, *cuando*) or less (typological contexts, Imperfect tense) linked to the chronological standard contexts of appearance of this copula. The Caribbean variety occupies an intermediate position, showing a higher reliance on concomitant factors (adverbs denoting change; *ya*) to induce *estar* but a lower incidence of the syntactic markers found in the most innovative zones. Finally, European and Rioplatense Spanish remain conservative, as shown by the absence of the factors taken as pointers to the innovative construction.

Our results confirm and refine the generalizations proposed by De Jonge (1993a, 1993b) and Malaver (2009) regarding the distribution of *estar*. Specifically, the Bayesian model confirms that *cuando* (when) clauses are a strong predictor for the appearance of *estar* in innovative varieties, marking what De Jonge termed neutral contexts. Furthermore, our analysis confirms the crucial role of the Imperfect tense (*estaba*) in *estar* with adjectives of age. The massive shift toward the Imperfect tense in the Andean and Mexican varieties distinguishes them from conservative varieties, where *estar* with age adjectives appears predominantly in the Present tense. The prevalence of the Imperfect tense suggests that *estar* is entering typological contexts thereby signaling a more advanced stage of grammaticalization in these areas. Conversely, other factors such as grammatical person or temporal adverbs expressing change (e.g., *ya*) do not appear to be statistically significant drivers of regional variation.

From a theoretical standpoint, these findings challenge the consideration of subjectivity or experiential meaning as uniform basis for the explanation for innovative *estar*. While recent proposals (Moreno-Quibén 2022; Gumiel-Molina, Moreno-Quibén and Pérez-Jiménez 2023, 2024) successfully attribute the use of *estar* with evaluative adjectives to a process of argument augmentation (involving a dative experiencer), this hypothesis faces difficulties with age adjectives, as claimed in section 2. Our data includes objective, typological predications where *estar* is used without an evaluative or scalar

nuance. In these instances, speaker disagreement would be factual, not “faultless”, undermining the claim that *estar* necessarily encodes subjective judgments in all innovative contexts. Note that in the following example, one of the speakers is right and the other is necessarily wrong, contrary to the “faultless disagreement” reading typical of subjective predications (Lasersohn 2005, 2009): Speaker A: *La bebé tenía tres meses, estaba chiquita* ‘The baby was.ESTAR 3 month old. She was small’—Speaker B: No, no lo estaba ‘No, she was.ESTAR not’. These kinds of examples do not seem to be evaluative judgments (scalar variation is not possible).

Consequently, the data support the view that the innovative use of *estar* with age adjectives implies a neutralization with the copula *ser* (as part of a diachronic process of extension, where *estar* progressively invades the semantic domain of *ser*, as claimed in the literature). In the most innovative varieties, this displacement indicates a reorganization of the copular system wherein *estar* assumes general descriptive functions, even in contexts of between-individual comparison. The fact that the “argument augmentation” hypothesis explains evaluative adjectives but fails to account for the objective use of age adjectives suggests that the innovative use of *estar* is not a single, homogeneous phenomenon. Rather, it appears to be a constellation of convergent phenomena dependent on the specific lexical-syntactic class of the adjective (as Escribano 2024 also notes to account for the distribution of standard and innovative *estar* with individual vs. eventive subjects with evaluative predicates).

In our opinion, future research should focus on the following lines. Investigating why age adjectives, specifically, dominate the innovative pattern and allow for objective readings that other adjectival classes (such as aesthetic adjectives, dimensional adjectives or property adjectives) do not; further exploring how the process of argument augmentation interacts with the broader grammaticalization of *estar*;<sup>10</sup> analyzing the coexistence of *ser* and *estar* in innovative regions. The data shows that, despite the expansion of *estar*, *ser* has not disappeared, suggesting a context of stable variation rather than total replacement. In summary, to understand this paradigm in its entirety, theoretical models must account for both the distinct lexical properties of adjectival classes and the varying stages of grammaticalization of the copula across the Spanish-speaking world.

## References

---

<sup>10</sup> It remains to be determined whether the “objective” use of *estar* with age adjectives represents a distinct path of change or an advanced stage of the same phenomenon affecting evaluative adjectives

- Aguilar-Sánchez, Jorge. 2012. «Formal Instruction and Language Contact in Language Variation: The Case of *ser* and *estar* + Adjective in the Spanish of Limón, Costa Rica». In *Selected Proceedings of the 14th Hispanic Linguistics Symposium*. Edited by Kimberly Geeslin and Manuel Díaz-Campos, 9-25. Somerville, Cascadilla Proceedings Project.
- Alfaraz, Gabriela G. 2012. «The Status of the Extension of *estar* in Cuban Spanish». *Studies in Spanish and Lusophone Linguistics* 5: 4–23. <https://doi.org/10.1515/shll-2012-1118>.
- Brown, Esther L., and Mayra 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 Kimberly Geeslin and Manuel Díaz-Campos, 61-74. Somerville, Cascadilla Proceedings Project, 61–74.
- Cortés-Torres, Mayra. 2004. «¿Ser o estar? La variación lingüística y social de *estar* más adjetivo en el español de Cuernavaca, Mexico». *Hispania* 87: 788–795.
- De Jonge, Robert. 1987. «*Estar* comes of age», in F. Beukema and P. Coopmans, eds., *Linguistics in the Netherlands 1987*, 101-110. Foris, Dordrecht.
- De Jonge, Robert. 1990. «Sprekersstrategieën en taalverandering: ser en estar in leeftijdsuitdrukkingen». Doctoral dissertation, Leiden University.
- De Jonge, Robert. 1991a. «Verschuivingen in het gebruik van *ser* en *estar* in Latijns-Amerika», in C. van Esch, and M. Steenmeijer, eds., *Spaans in onderwijs onderzoek en bedrijfsleven* 2, 50-61. KUN, Nijmegen1.
- De Jonge, Robert. 1991b. «La cosa (no) es como está », in C. Hernandez, G.P. Granda, C. Hoyos, V. Fernandez, D. Dietrick and Y. Caballera, eds., *El español de América I, Actas del III Congreso Internacional de El Español de América*, 495-504. Junta de Castilla y León, Salamanca.
- De Jonge, Robert. 1993a. «(Dis)continuity in language change: *ser* and *estar* + age in Latin American Spanish». In *Linguistics in the Netherlands 1993*, eds. F. Drijkoningen y K. Hengeveld, 69-80. Amsterdam–Philadelphia, John Benjamins. <https://doi.org/10.1075/avt.10.09jon>.
- De Jonge, Robert. 1993b. «Pragmatismo y gramaticalización en el cambio lingüístico: *ser* y *estar* en expresiones de edad». *Nueva Revista de Filología Hispánica* 41(1): 99–126. <https://doi.org/10.24201/nrfh.v41i1.925>.
- Díaz-Campos, Manuel, and Kimberly L. 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. <https://doi.org/10.1075/sic.8.1.04dia>.
- Escandell-Vidal, Victoria, and Manuel Leonetti. 2016. «*Estar* joven a los dos lados del Atlántico». In *Geométrica explosión. Estudios de lengua y literatura en homenaje a René Lenarduzzi*, eds. M. E. Sainz González, I. Solís García, F. del Barrio de la Rosa e I. Arroyo Hernández, 65–77. Venecia: Edizioni Ca' Foscari. <https://doi.org/10.14277/6969-068-6/RiB-1-4>.
- Escandell-Vidal, Victoria. 2018. «Evidential commitment and feature mismatch in Spanish *estar* constructions». *Journal of Pragmatics* 128: 102–115. <https://doi.org/10.1016/j.pragma.2017.10.004>
- Escandell-Vidal, Victoria. 2023. «*Estar*+ILP. Testing the experiential commitment», *Spanish in Context*, 20(2): 257-281 .
- Escribano, Gonzalo. 2024. «Events and copulas: An approach to a phenomenon of variation across Spanish dialects». *Isogloss. Open Journal of Romance Linguistics* 10(5), art. 7. <https://revistes.uab.cat/isogloss/article/view/v10-n5-escribano>.
- Falk, Johan. 1979. *Ser y estar con atributos adjetivales*. Upsala, Alqvist and Wiksell.

- García-Márkina, Yekaterina. 2013. «Recherches sur l'opposition entre ser et estar en espagnol. Historique de la question, et application à l'étude des variations dans leurs emplois en espagnol spontané contemporain au Mexique». Doctoral dissertation, l'Université Sorbonne Nouvelle, Paris, France.
- Gelman, Andrew, and Jennifer Hill. 2021. *Data analysis using regression and multilevel/hierarchical models*. Cambridge, Cambridge Univ. Press.
- Goodrich Ben, Gabry Jonah, Imad Ali and Sam Brilleman. 2025. *rstanarm: Bayesian applied regression modeling via Stan*. R package version 2.32.2. <https://mc-stan.org/rstanarm>.
- Gumiel-Molina, Silvia, Norberto Moreno-Quibén and Isabel Pérez-Jiménez. 2015. «A syntactic approach to the relative/absolute distinction: The case of adjectives in copular sentences in Spanish». *Natural Language and Linguistic Theory* 33: 955–1001.
- Gumiel-Molina, Silvia, Norberto Moreno-Quibén and Isabel Pérez-Jiménez. 2020. «Variación dialectal en las oraciones copulativas en español: los adjetivos de edad y tamaño». In *Elementos de variación y diversidad: la palabra*. Edited by E. Hernández and P. Martín Butragueño, 219-244. México, Ediciones del Colegio de México/CSIC.
- Gumiel-Molina, Silvia, Norberto Moreno-Quibén and Isabel Pérez-Jiménez. 2023. «Perspectivized *estar*-sentences with aesthetic adjectives across American Spanish varieties». *Spanish in Context*. 20(2): 313-342. <https://doi.org/10.1075/sic.00086.gum>
- Gumiel-Molina, Silvia, Norberto Moreno-Quibén and Isabel Pérez-Jiménez. 2024. «Lexical–syntactic classes of adjectives in copular sentences across Spanish varieties: The innovative use of *estar*». *Languages* 9(1). <https://doi.org/10.3390/languages9010020>.
- Gutiérrez, Manuel. 1992. «The extension of *Estar*: A linguistic change in progress in the Spanish of Morelia, Mexico». *Hispanic Linguistics* 5: 109–141.
- Lasersohn, Peter. 2005. «Context dependence, disagreement and predicates of personal taste». *Linguistics and Philosophy* 28: 643–86. <https://doi.org/10.1007/s10988-005-0596-x>.
- Lasersohn, Peter. 2009. «Relative truth, speaker commitment, and control of implicit arguments». *Synthese* 166: 359–74. <https://doi.org/10.1007/s11229-007-9280-8>.
- Malaver, Irania. 2009. *Variación dialectal y sociolingüística de ser y estar con adjetivos de edad*. Ph.D Thesis. Universidad de Alcalá.
- Malaver, Irania. 2022. «El español en Venezuela». In *Dialectología hispánica / The Routledge Handbook of Spanish Dialectology*, eds. Francisco Moreno Fernández and Rocío Caravedo, 394–405. London–New York, Routledge.
- McElreath, Richard. 2020. *Statistical rethinking: a Bayesian course with examples in R and Stan* Second edition. Boca Raton: Chapman & Hall/CRC.
- Moreno Fernández, Francisco. 2009. *La lengua española en su geografía*. Madrid: Arco Libros.
- Moreno Fernández, Francisco. 2019. *Variedades de la lengua española*. London–New York: Routledge.
- Moreno-Quibén, Norberto. 2022. «Las estructuras de predicación y la alternancia de cópula». Doctoral dissertation, Universidad de Alcalá, Madrid.
- Nicenboim, Bruno, and Shravan Vasishth. 2016. «Statistical methods for linguistic research: Foundational Ideas—Part II». *Language and Linguistics Compass* 10(11): 591–613. <https://doi.org/10.1111/lnc3.12207>.

- Orozco, Rafael. 2022. «El español en Colombia (Spanish in Colombia)». In *Dialectología hispánica / The Routledge Handbook of Spanish Dialectology*, eds. Francisco Moreno Fernández y Rocío Caravedo, 227–240. London–New York, Routledge.
- Pérez-Jiménez, Isabel, Silvia Gumiel-Molina and Norberto Moreno-Quibén. 2018. «Ser y estar en las lenguas romances ibéricas: las oraciones copulativas con atributo adjetival». *Revista Española De Lingüística* 48(1): 153–198. <http://doi.org/0.31810/RSEL.48.5>
- Piñango, María Mercedes, and Zuzanna Fuchs. 2023. «The real-time processing of *ser* and *estar* in Spanish». *Glossa: A Journal of General Linguistics* 8(1): 1–31. <https://doi.org/10.16995/glossa.9123>.
- PRESEEA (2014-): *Corpus del Proyecto para el estudio sociolingüístico del español de España y de América*. Alcalá de Henares: Universidad de Alcalá. [<http://preseea.uah.es>].
- R Core Team. 2025. *R: A language and environment for statistical computing*. Vienna: Foundation for Statistical Computing. <https://www.R-project.org/>.
- RAE-ASALE. 2009. *Nueva gramática de la lengua española*. Madrid, Espasa.
- RAE-ASALE. 2025. *Nueva gramática de la lengua española: actualización y ampliación*. Madrid, Espasa.
- Sánchez-Alonso, Sara, María Mercedes Piñango and Ashwini Deo. 2019. «Variability in *ser/estar* use across five Spanish dialects: An experimental investigation». Manuscript. Available on line: [https://www.researchgate.net/publication/335568928\\_Variability\\_in\\_serestar\\_Use\\_Across\\_Five\\_Spanish\\_Dialects\\_An\\_Experimental\\_Investigation](https://www.researchgate.net/publication/335568928_Variability_in_serestar_Use_Across_Five_Spanish_Dialects_An_Experimental_Investigation) (accessed on 29<sup>th</sup> December 2025).
- Sánchez-Alonso, Sara. 2018. *The Cognitive Sources of Language Change and Variation: Connecting Synchronic Variation and Dia-chronic Spanish Copula Use*. Ph.D. thesis, Yale University, New Haven, CT, USA.
- Silvagni, Federico. 2021. *La gramática de ser y estar*. Madrid: Arco/Libros.
- Sorensen, Tanner, Sven Hohenstein and Shravan Vasishth. 2016. Bayesian linear mixed models using Stan: A tutorial for psychologists, linguists, and cognitive scientists. *The Quantitative Methods for Psychology* 12(3): 175–200. <https://doi.org/10.20982/tqmp.12.3.p175>
- Vasishth, Shravan, and Bruno Nicenboim. 2016. Statistical Methods for Linguistic Research: Foundational Ideas – Part I. *Language and Linguistics Compass* 10(8): 349–369. <https://doi.org/10.1111/lnc3.12201>.
- Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., & Dunnington, D. (2021). *ggplot2: Create elegant data visualisations using the grammar of graphics* (R package).
- Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.