

Do bilinguals generalize *estar* more than monolinguals and what is the role of conceptual transfer?

International Journal of Bilingualism

1–32

© The Author(s) 2018

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/1367006918812175

journals.sagepub.com/home/ijb**Evangelia Adamou**

CNRS (French National Centre for Scientific Research), France

Stefano De Pascale

KU Leuven, Belgium

Yekaterina García-Márkina

University of Tours, France

Cristian Padure

Inalco (National Institute for Oriental Languages and Civilizations), France

Abstract

Aims and objectives/purpose/research questions: Among the questions that remain open is whether bilingualism leads to simplification of alternatives in language in order to reduce cognitive load. This hypothesis has been supported by evidence showing that bilinguals generalize the Spanish copula *estar* ‘to be’ faster than monolinguals. Yet, other studies found no such clear trend. While conceptual transfer could account for the conflicting evidence in the literature, its role has not been demonstrated. Our study aims to fill this gap by testing simplification in Spanish copula choice among bilinguals and, in particular, the role of transfer.

Design/methodology/approach: We used a contextualized copula choice task, comprising 28 sentences.

Data and analysis: Sixty Romani–Spanish bilinguals from Mexico responded to the questionnaire in both Spanish and Romani. A control group of 62 Mexican Spanish monolinguals responded in Spanish. We constructed generalized linear mixed-effects models to analyse the results.

Findings/conclusions: Analysis of the results reveals greater extension of *estar* among bilinguals for individual-level predicates as well as for traits not susceptible to change. Comparison of the responses of bilinguals (in Romani and in Spanish) and of Spanish monolinguals indicates that Romani could be reinforcing the generalization of *estar* in the Spanish responses of bilinguals.

Originality: To our knowledge, this is the first study to examine copula choice in bilingual mode. In addition, it brings evidence from an under-researched community with little normative pressure.

Corresponding author:

Evangelia Adamou, CNRS, LACITO, 7 rue Guy Moquet, Villejuif 94801, France.

Email: evangelia.adamou@cnrs.fr

Significance/implications: Our study shows that conceptual transfer may be driving the extension of *estar* among bilinguals.

Keywords

Bilingualism, copulas, simplification, Spanish, Romani

Introduction

Bilingualism poses some of the greatest challenges in accounting for the way the human mind works and how this can affect language in the long run. Among the questions that remain open is whether bilingualism leads to simplification of alternatives in language in order to reduce cognitive load. Silva-Corvalán (1986) formulated this hypothesis in a pioneering article based on the study of the Spanish copulas *ser* ‘to be’ and *estar* ‘to be’. The author showed that English–Spanish bilinguals with higher degrees of exposure to English generalize *estar* ‘to be’ to contexts previously covered by *ser* ‘to be’, thus leading to the simplification of some selectional restrictions. Geeslin and Guijarro-Fuentes (2008), however, found no clear trend in a large-scale study of four bilingual communities of the Iberian Peninsula. While conceptual transfer between the languages in contact could account for the conflicting evidence in the literature, its role has not been demonstrated yet. Our study therefore aims to fill this gap by testing simplification in Spanish copula choice among bilinguals and, in particular, by assessing the role of conceptual transfer. In accordance with the above-mentioned studies, we define simplification of alternatives as the loss of some constraints in the selection of alternatives and the subsequent increase of rates of one of the alternatives. Evidence comes from an under-investigated bilingual population of heritage Romani speakers residing in Mexico for the past 150 years. In the remainder of this introduction, we present a brief overview of research on conceptual transfer and simplification processes, in the ‘Conceptual transfer and simplification of alternatives among bilinguals’ section, on copula choice in Spanish, in the ‘Spanish copula variation’ section, and on Mexican Romani, in the section titled ‘Background on Romani’.

Conceptual transfer and simplification of alternatives among bilinguals

When two languages in contact have different conceptual representations, which are encoded by distinct linguistic means, conceptual transfer is likely to occur from one language to another. Following Schmid and Köpke (2017), we view transfer as an online process that can potentially also modify underlying representations at the level of the individual speaker. Transfer at the level of the individual may eventually lead to language change and convergence between the two languages when systematically observed among several members of a given community, sometimes over generations of speakers. The direction of transfer is traditionally thought to take place from the first language (L1) to the second language (L2) (based on the chronological order of acquisition), or from the socially dominant language to the heritage or the minority language. However, recent investigations reveal that cross-language interactions are most likely bidirectional (see Jarvis & Pavlenko, 2008; Kroll et al., 2015; Schmid & Köpke, 2017). As we consider transfer to be a constantly dynamic, bidirectional process, we do not distinguish between transient and permanent effects as suggested for example by Grosjean (2011), who restricts the term ‘transfer’ to permanent effects and ‘interference’ to ephemeral effects. Multiple factors may be at play in determining the degree and direction of transfer: the age at which the speaker becomes bilingual; proficiency in the two languages; frequency of exposure to the L1 and the L2 and degree of coactivation of the L1 and the L2; typological similarity between the two languages and linguistic phenomena under discussion as well as types of linguistic

phenomena (see among others Bylund, 2009; Gollan et al., 2008; Green & Abutalebi, 2013; Kroll & Dussias, 2013; Montrul, 2006; Sorace & Serratrice, 2009).

In this paper, we explore the role of conceptual transfer in Spanish copula choice among bilinguals. On the one hand, it is well known that English L2 learners of Spanish encounter difficulties in the use of the Spanish copulas *ser* 'to be' and *estar* 'to be', which correspond to two conceptual representations, whereas in English they have a single conceptual representation of *being*, expressed by the copula *to be*. Such late bilinguals, who learn Spanish in the classroom, first generalize *ser* 'to be' before acquiring *estar* 'to be' (e.g. Geeslin, 2001). In contrast, based on data from three generations of heritage speakers of Spanish in Los Angeles, USA, Silva-Corvalán (1986, 1994) showed that the younger speakers generalize *estar*. Gutiérrez (1994) then compared the older generation of immigrants examined by Silva-Corvalán with Spanish monolinguals from Michoacán, Mexico, and noted the same trend. Gutiérrez's results suggest that the generalization of *estar* started in Mexico and was accelerated in the contact setting with English in the USA. Subsequent research confirmed the generalization of *estar* in other monolingual Spanish-speaking communities in the Americas (see Díaz-Campos & Geeslin, 2011 for Venezuela; García-Márkina, 2013 for Mexico).

Silva-Corvalán (1986: 588, 1994) hypothesizes that the faster generalization of *estar* among bilinguals stems from general cognitive factors, which favour simplification of alternatives. However, Ortiz López (2000: 111) provides counter-evidence showing that Spanish (quasi)-monolinguals in Puerto Rico use *estar* more innovatively than Spanish–English bilinguals do. In addition, Geeslin and Guijarro-Fuentes (2008) find no evidence for the generalization of *estar* among the various bilingual communities of the Iberian Peninsula. Moreover, Adamou (2013) shows that Mexican Romani has been rendered more complex by copying the Spanish copula variation and concludes that convergence of the two grammars is ultimately simplifying matters for bilinguals.

Geeslin and Guijarro-Fuentes (2008: 376) acknowledge that the role of copula choice in the contact languages needs to be further examined. Similarly, Silva-Corvalán (1986) considers that although her study does not provide any evidence that conceptual transfer from English is driving the process of simplification of the Spanish copula choice, transfer is a potential explanatory factor, in particular when the contact language also exhibits copula variation (Silva-Corvalán, 1994).

In the present study, we set out to test the role of transfer by examining Spanish copula choice in bilingual mode, that is, when a speaker uses the two languages in the same setting and with the same interlocutors (see Soares & Grosjean, 1984 for an early study of the effects of mode in bilingual research). One important reason for opting for the bilingual mode is that it is closer to the in-group communicational habits of the community under study. Indeed, a growing number of researchers argue that it is important to take into consideration the real-life communicational habits of the individuals in their community when investigating bilingualism and its effects (Adamou & Shen, 2017; Green & Abutalebi, 2013). The second motivation behind the choice of the bilingual mode is that it can offer a window to the mechanism of long-term, contact-induced language change through observation of real-time, albeit ephemeral, cross-language priming. For example, Torres Cacoullos and Travis (2018) show that priming, a general mechanism in language production that leads to the repetition of a previously mentioned structure, is the best predictor for structural convergence. However, authors call for a cautious account of this effect given that cross-language priming seems to be less strong than within-language priming.

Another originality of the current study is its focus on an under-investigated population characterized by low normative pressure, involving heritage speakers of Romani (Indic) residing in Veracruz, Mexico, who are bilingual with the majority language, Mexican Spanish (Romance). We compare the results to copula choice in a monolingual control group of Mexican Spanish speakers. Although the study by Silva-Corvalán (1986) involves Spanish as a heritage language, she suggests that the same acceleration should be observed in a majority language provided there is extended language contact (Silva-Corvalán, 1986: 604), as is the case for Mexican Spanish in this study.

Spanish copula variation

The distribution of *ser* and *estar* was first examined through a binary opposition where *ser* is associated to permanent and essential properties and *estar* to temporary and accidental properties. However, in the past decades, researchers have proposed a new set of syntactic, semantic and pragmatic parameters. In particular, several authors account for the *ser* and *estar* variation through aspect (perfectivity, resultant states or bounded vs unbounded states) (e.g. Camacho, 1997; Luján, 1981; Marín Gálvez, 2004; Schmitt, 1992). Others elaborate detailed classifications of adjectives that accept *ser* and/or *estar* (see Navas Ruiz, 1963; Vaño Cerdá, 1982). Moreover, a number of researchers put forward the importance of predicate type. They show that *ser* is associated to individual-level properties, which are not limited in time and apply to an individual as a whole, as in *Elena es simpática* 'Elena is (*ser*) nice'. In contrast, *estar* is associated to stage-level properties, which are more limited in time, as in *Hoy, Elena está enferma* 'Today, Elena is (*estar*) sick' (e.g. Arche, 2006; Escandell-Vidal & Leonetti, 2002; Fernández Leborans, 1995; Lema, 1992). In addition, Delbecque (2000) focuses on the semantic motivation (deictic vs non-deictic predication), and Porroche (1988), among others, on the subject type (animate vs inanimate). Other accounts of the variation rely on the speaker's point of view. A well-known factor is the distinction between class and individual frames of reference, most notably in Falk (1979). In a class frame of reference, the referent is compared to a group of referents, as in *Juan es alto* 'John is (*ser*) tall'; in an individual frame of reference, the referent is compared to itself at some other point in time, as in *Juan está alto* 'John is (*estar*) tall', intended meaning 'John has grown tall'. Other distinctions include contextual and discourse factors (Clements, 1988), sensorial experience (Maienborn, 2005), and the expression of evidential interpretation (Camacho, 2015; Escandell-Vidal, 2017; García-Márkina, 2013).

Following Silva-Corvalán (1986), non-linguistic variables are also taken into account such as gender, age, socio-economic status, level of instruction, language knowledge, variety of Spanish, and stylistic factors (see among others Cortés-Torres, 2004; De Jonge, 1993; Díaz-Campos & Geeslin, 2011; Geeslin & Guijarro-Fuentes, 2008; Gutiérrez, 1994; Malaver, 2012; Ortiz López, 2000). Recently, in a usage-based approach, Brown and Cortés-Torres (2012) stress the importance of the constructions combining the copula and the adjective as a whole.

Background on Romani

Estimates of Roma numbers in the Americas range from 1.5 to 3.5 million¹ and yet there is very little research about their migrations, way of life and social organization. Most of the Roma probably migrated to the Americas along various routes in the nineteenth century as part of the more general European migration. Some Romani communities had already settled in countries such as Brazil and Argentina following earlier migrations from Spain and Portugal starting in the sixteenth century. Historical documents also mention the presence of *gitanos* in eighteenth-century Mexico. Another important wave of migration took place after the Second World War and mobility to various countries in Northern, Central and South America continue today.

In Mexico, Roma have settled in several cities, mainly in Mexico City, but also in Veracruz, Guadalajara, Oaxaca, Tuxtla Gutiérrez and Puebla. The data presented in this paper were collected in 2016 and 2017 in the small community of La Rinconada, in the State of Veracruz, Mexico (see map in Figure 1). Most Roma in the community work in the car trade. They live in mixed neighbourhoods and intermarry with other Roma living in Mexico or with outsiders. Men, women and children are bilingual in Romani and Mexican Spanish. The Romani variety spoken in Veracruz shares several features with the south-eastern Romani dialects of Europe and more specifically the so-called Vlax dialects (dialect classification in Matras, 2005), similar to the Romani variety spoken in Oaxaca (Adamou, 2013).



Figure 1. Map of Mexico. The Romani–Spanish bilinguals reside in the locality of La Rinconada in the State of Veracruz.

Adamou (2013) argues that, under the influence of the Spanish copulas, Mexican Romani speakers developed a distinction between attributive predications using the copula *si* ‘to be’, as in (1a), or the third person subject clitic pronouns in *l-*, as in (1b), whereas Romani speakers from Europe only use the copula (Elsik & Matras, 2006; Matras, 2002). The result is the complexification of the Romani grammar (Adamou, 2013).

(1a) le ʃave muʒa bibiake si barbale
 DEF.PL children POSS.1SG aunt.DAT be.3PL rich
 ‘My aunt’s children are rich.’ (Adamou 2013: 1085)²

(1b) o raklo=**lo** felis
 DEF.M boy-3SG.M happy
 ‘The boy is happy.’ (Adamou 2013: 1075)

The influence of Spanish *estar* on the Romani grammar is not restricted to the attributive clauses; the subject clitics have equally replicated the uses of *estar* in locative predications and in constructions with participles (Adamou, 2013: 1075–1076).

The clitic pronouns inflect for gender and number, *lo* masculine singular, *la* feminine singular, and *le* plural. They are dubbed *clitics* because their position in the clause is not fixed, but they always need to attach (*encliticize*) to other words. Subject clitic pronouns in *l-* are an archaism in Romani and a new set of subject pronouns has replaced them in all Romani dialects (Matras, 2002: 102, 111). The use of the *l-* clitics in attributive predications is not documented in Europe, but it is reported for Romani as spoken in Bogota, Colombia (pilot study in Acuña & Adamou, 2013). Padure, De Pascale and Adamou (in press) encounter the same phenomenon in the Romani

variety spoken in Veracruz, Mexico, based on the analysis of a 15-hour-long corpus of interviews from 19 speakers. Examples in (2) illustrate the variation between the copula, in (2a), and the clitics, in (2b).

Although analysis of the conversational corpus confirms the variation between the clitics and the copula, it provides a limited number of occurrences that cannot improve understanding of the

(2a) vo motholas ke **si** tʃoɾo
 3SG.M screamed.3SG that be.3SG poor
 ‘He screamed that he is poor.’ (SP28; 1:15:42)

(2b) o them kathe tʃoɾo=**lo**
 DEF.SG.M country here poor-3SG.M
 ‘The country here is poor.’ (SP12; 0:45)

variation, that is, 50 occurrences of [copula + adjective] and 66 occurrences of [clitic + adjective]. In order to explore this variation and the equivalence between Mexican Romani and Spanish more systematically, we conducted a quantitative study based on a contextualized copula choice task. The next section presents the method, the following section the results, and the penultimate section contains a discussion of the results.

Method

Goals and predictions

As the previous review shows, there is ongoing discussion about whether bilingualism leads to simplification of alternatives in Spanish copula choice (e.g. Geeslin & Guijarro-Fuentes, 2008; Silva-Corvalán, 1986). Although the aforementioned studies manipulated a number of linguistic and extra-linguistic variables, they did not directly test the role of conceptual transfer. The present study seeks to tease apart these factors and assess their relative weight. More specifically, the study is guided by two research questions:

Research question 1: Is there simplification of alternatives in Spanish copula choice among bilinguals as compared to monolinguals?

Predictions: We predict simplification of alternatives and increase of *estar* following the results from the USA in Silva-Corvalán (1986). In particular, we expect the linguistic variable ‘frame of reference’ (class vs individual) to drive the process, followed by ‘susceptibility to change’ (change vs no change) (Silva-Corvalán, 1986: 595).³ Lack of generalization of *estar* would need to be examined in the light of the complex set of linguistic and sociolinguistic factors that determine copula choice more generally (Geeslin & Guijarro-Fuentes, 2008).

Research question 2: Is simplification of alternatives in Spanish copula choice among bilinguals due to transfer from Romani?

Predictions: Both Silva-Corvalán (1986) and Geeslin and Guijarro-Fuentes (2008) consider conceptual transfer from the contact language to be a potential factor of the observed copula choice in Spanish. In our study, based on the contact-induced effects that took place at the level of the

Romani copulas (Adamou, 2013), we expect conceptual transfer to occur at least from Spanish to Romani. In addition, given the design of the task where Romani speakers translate the Spanish response into Romani, we expect some transient, online effects to show, in particular because of cross-language priming. In the light of literature reporting bidirectional transfer (see Kroll et al., 2015; Schmid & Köpke, 2017), we further predict transfer from Romani to Spanish.

Participants

Sixty Romani–Spanish bilinguals from Veracruz, Mexico participated in this study (48 males; women are under-represented in the study due to cultural norms restricting co-ed sessions). Thirty-two participants declared being early simultaneous bilinguals (2L1, i.e. before the age of 3), 27 being early sequential bilinguals (L1 Romani–L2 Spanish), and one was a late bilingual having acquired Romani after the age of 18. Fifty-seven participants had attended at most primary school and three high school. Age of participants ranged from 17 to 90 (29 participants were 17–29 years old, 19 participants were 30–59 years old, and 12 participants were 60–90 years old). Participants were all residents of La Rinconada community, in the State of Veracruz, Mexico at the time of the recording. They all had a similar socio-economic status, working in the car trade or doing housework. Two participants were attending high school at the time of the study. Participants gave written consent and received no financial compensation for their participation in the study, but the interviewer followed local custom and organized a celebration dinner to which all community members were invited.

In order for our study to have a solid, comparative basis beyond differences with an idealized monolingual norm, we tested a control group of 62 Mexican Spanish monolinguals. As it is challenging to match the monolingual and the bilingual groups on all extra-linguistic variables, we focused on those that were identified as significant in most other studies on Mexican Spanish, namely age and education. The monolingual participants resided in Mexico City (Topilejo).⁴ They matched the bilingual group for age, socio-economic status, and, as much as possible, for education; indeed, the monolingual group had relatively higher levels of education than the bilingual group, as 33 participants had received secondary or technological education and three declared having attended college.⁵ We contacted participants through religious networks, social medical services or through social networks within the neighbourhood. All responded to the questionnaire on a voluntary basis and received no compensation for their participation.

Materials

We used the contextualized copula choice task, which Geeslin and Guijarro-Fuentes (2008) developed in Spanish. We opted for this task for two main reasons. First, because it is possible to compare the Mexican results to those of the studies that used the same task. Second, because this task offers the possibility to investigate the main linguistic variables that determine copula choice in comparable contexts – something that is very difficult to obtain with the analysis of spontaneous data.

The contextualized copula choice task contains 28 items introduced by a paragraph-long context connected in a way that the entire task forms a coherent story. Participants have three options: they can opt for a sentence with *ser*, a sentence with *estar*, or indicate that they like both *ser* and *estar* in this context. To exemplify the task, we present an excerpt (Guijarro-Fuentes & Geeslin, 2006: 69):

Paula y Raúl salen del apartamento y van al restaurante. Comen allí frecuentemente y la gente que trabaja en el restaurante siempre los trata bien. Esta vez, Raúl pidió algo nuevo y Paula quiere saber qué piensa Raúl de la comida. Paula: Raúl, ¿te gusta la comida?

‘Paula and Raúl leave the apartment and go to a local restaurant. They eat there frequently and the people who work there are always very nice. This time, Raúl has ordered something new on the menu and Paula is curious about what Raúl thinks of the food. Paula: Raúl, do you like your food?’

A. Raúl: *Sí, la cena es buena.*

A. ‘Raúl: Yes, dinner is good (ser).’

B. Raúl: *Sí, la cena está buena.*

B. ‘Raúl: Yes, dinner is good (estar).’

___ **Prefiero la frase A.**

‘___ I prefer sentence A.’

___ **Prefiero la frase B.**

‘___ I prefer sentence B.’

___ **Prefiero A y B.**

‘___ I like both A and B.’

Each discourse context and related items were designed to control for the main linguistic variables that determine copula choice: ‘predicate type’, individual-level property vs stage-level property; ‘susceptibility to change’, susceptible to change vs not susceptible to change; ‘animacy’, animate vs inanimate; ‘frame of reference’, class frame vs individual frame; ‘experience with the referent’, ongoing versus immediate; ‘adjective class’, grouping adjectives into ten classes depending on their properties such as age or size and status; and ‘copulas allowed’, *ser* only versus *estar* only versus both, based on the adjectives that regularly appear with both copulas in usage and those that preferably combine with one of the two copulas. The additional use of this task for the study of Romani copula choice required the consideration of new linguistic variables that the Spanish version of the task was not designed to explore. In particular, as variation in Romani is only possible in third person affirmative clauses, we added two structural variables in the analysis, that is, clause type (affirmative vs negative) and person (third person vs first and second person). As a result, the study of the Romani–Spanish equivalence relied on a reduced and unbalanced questionnaire.⁶ This had an effect on the statistical analyses as discussed in detail in the ‘Analysis’ section.

Procedure

The bilingual participants were tested in their homes. A researcher who is a native Romani speaker of a closely related Vlax dialect from Europe conducted the testing. After giving their oral consent, participants listened to 28 clauses in Mexican Spanish introduced by a paragraph-long context. For each clause, they were asked to choose between the copulas *ser* and *estar* or to indicate when both were applicable. The participants were then immediately asked (in Romani) to translate the target clauses into Romani. In total, each participant responded to 56 questions. The monolingual participants were tested in Spanish in their congregation’s church, the clinic’s waiting room, or the homes of the participants.

Analysis

In the current study, we used mixed-effects models in order to account for variability among both subjects and items and take into account interaction effects between the linguistic variables. Statistical analyses were performed using the open source statistical software R (R Core Team, 2015), and in particular the package *lme4* (Bates et al., 2015) for the *glmer* function.⁷

More specifically, we carried out statistical analyses in order to compare Spanish copula choice and the predictors that might affect this choice in the bilingual and monolingual groups. The initial dataset includes all the experimental items for which this variation is possible, that is, for all

Table 1. The fixed effects used in the statistical analyses of the Spanish responses for the bilingual and monolingual groups.

	Fixed effects	Levels
Language (<i>bilingual or monolingual</i>)	copulas allowed in Spanish (abbr. copula choice allowed)	<i>both, ser, estar</i>
	animacy	<i>animate, inanimate</i>
	experience with referent	<i>immediate, ongoing</i>
	predicate type	<i>stage, individual</i>
	susceptibility to change (abbr. change)	<i>change, no change</i>
	clause type	<i>affirmative, negative</i>

persons and numbers in the verb paradigm and for both affirmative and negative sentences. From this initial dataset we discarded experimental items in which respondents deemed both copulas appropriate – approximately 6.7% of data points by the monolingual group and 1.7% by the bilingual group [final token size = 3081 points].

We built generalized linear mixed-effects models with ‘Spanish copula’ as a response variable with two levels, *ser* and *estar*. Random intercepts were included for ‘Participant’ and ‘Experimental item’. Table 1 summarizes the different fixed effects used in this study, that is, ‘copula choice allowed’ (both vs *ser* vs *estar*), ‘animacy’ (animate vs inanimate), ‘experience with referent’ (immediate vs ongoing), ‘predicate type’ (stage vs individual), ‘susceptibility to change’ (change vs no change), all coded as in Geeslin and Guijarro-Fuentes (2008),⁸ to which we added ‘clause type’ (affirmative vs negative), which was of interest to us because of the fact that variation in Romani is only possible in the affirmative clauses. We did not include ‘adjective class’ in the regression models despite its potential relevance, as it has 10 different levels and integrating it in the model would lead to serious issues in the coefficient calculations. The results section investigating the Romani–Spanish equivalence contains descriptive statistics on the relation between ‘adjective class’ and copula choice (note that all adjective classes were represented by at least one item).

Given our interest in uncovering differential predictor strength for our two groups of participants, we made those fixed effects, and their two-way interactions, interact in turn with the variable ‘language’, potentially allowing for significant three-way interactions.⁹

Following the recommendations in Barr et al. (2013), we opted for a stepwise backward model selection procedure on the maximal models with respect to the fixed-effects structure (i.e. on all predictors and their three-way interaction), and retained both ‘participant’ ($p < 0.001$) and ‘experimental item’ ($p < 0.001$) as random effects.

We then aimed to investigate the extent to which observed differences between monolinguals and bilinguals can be explained in terms of conceptual transfer, either from Romani to Spanish or from Spanish to Romani. However, the copula alternation in Romani takes place in a more limited number of contexts compared to Spanish. Since the copula alternation in the bilinguals’ Romani dialect only appears in third person affirmative clauses, we restricted the dataset used in the previous section to this type of sentence (so-called variable context). In consequence, responses to items 1, 3, 8, 9, 10, 12, 13, 24, 25, 27, 28 of the Geeslin and Guijarro-Fuentes (2008) design were not taken into account in the datasets of both bilinguals and monolinguals (these items were either negative or/and first and second person). Similar to the previous analyses, we discarded experimental items for which respondents considered both Spanish copulas appropriate (in the monolingual data and the Spanish data of the bilinguals), but also those in which the Romani translation did not feature a copula at all (in the Romani data of the bilinguals). This means that the bilinguals’ dataset was reduced to 858 data points (approximately 58% of the data points used for the previous

analysis), and that 978 data points were retained for the monolingual speakers (approximately 61% of the data used in the analysis presented in section 3). We then constructed generalized linear mixed-effects models (see Table 3 in Appendix B for details). However, because of poor model performance, we chose not to base the analysis on the fitted probabilities of the models but only use the significant interactions to guide the qualitative analyses.

Results

We start by presenting the results relevant to the first research question, based on the analysis of the Spanish responses for both groups; see ‘The Spanish responses of the bilingual and monolingual groups’ section. We then look at the results regarding the second research question, based on the analysis of the Romani and Spanish responses of the bilinguals and the equivalent Spanish responses of the monolinguals; see ‘The Romani and Spanish responses among the bilinguals and the Spanish responses among the monolinguals’ section.

The Spanish responses of the bilingual and monolingual groups

In this sub-section, we address the first research question: is there simplification of alternatives in Spanish copula choice among bilinguals as compared to monolinguals? First, when looking at the descriptive statistics of the rates of selection, we note the generalization of *estar* among bilinguals, with 55.77% of *estar* choice as opposed to 45.85% among monolinguals. We then turn to the results of the generalized linear mixed-effects regression on the bilinguals’ and monolinguals’ choice of Spanish copula that confront the rates of selection to the linguistic conditioning. Analysis of the results reveals a two-way interaction effect of ‘susceptibility to change’ and ‘animacy’ ($\chi^2 = 7.6285$, $df = 1$, $p < 0.01$) as well as ‘predicate type’ and ‘clause type’ ($\chi^2 = 10.2944$, $df = 1$, $p < 0.01$) across both groups of participants. This means that the joint influence of these constraints on the copula alternation applies in an undifferentiated way for both bilinguals and monolinguals. In this respect, no three-way interactions (involving the ‘language’ predictor) were retained by the model selection procedure. In significant two-way interactions with ‘language’, only ‘predicate type’ ($\chi^2 = 18.0996$, $df = 1$, $p < 0.001$) and ‘change’ ($\chi^2 = 20.8762$, $df = 1$, $p < 0.001$) were retained, highlighting the differential impact of these constraints on the choice of Spanish copula in the two groups of participants. Furthermore, ‘copulas allowed in Spanish’ ($\chi^2 = 19.1655$, $df = 2$, $p < 0.001$) turned out to be a significant main effect. To inspect in more detail the influence of all these interactions, we resort to the following plots of the fitted probability for the success outcome in the model (i.e. *estar*). The model nearly attains good predictive power ($C = 0.796$) and classification accuracy is well above chance level (73% as compared to 53% for always choosing the most frequent response level, i.e. *estar*). We will now turn to the presentation of the interactions in detail (also see Appendix Table A.1).

The first two-way interaction is between ‘susceptibility to change’ and ‘animacy’ (see Figure 2). The fact that this interaction does not reach significance level in a three-way interaction with the predictor ‘language’ shows that bilinguals and monolinguals share the combined impact of those two sentence properties on the choice of copula. It is clear that the trait’s susceptibility to change does not matter when the referent is inanimate: in both cases, participants variably alternate between the use of *estar* and *ser*. We do note, however, a slight preference for *estar*, especially for traits that are susceptible to change, as in *La cena está buena* ‘Dinner is (*estar*) good’. In the case of an animate referent combining with a trait that is not susceptible to change, participants overwhelmingly resort to *ser* (100-14=86%), as in *Ahora ella es católica también* ‘Now she is (*ser*) Catholic too’. In contrast, both groups alternate between *ser* and *estar* when the referent is animate and the trait is susceptible to change, for example, *Ahora es/está enojado* ‘Now he is (*ser/estar*) angry’.

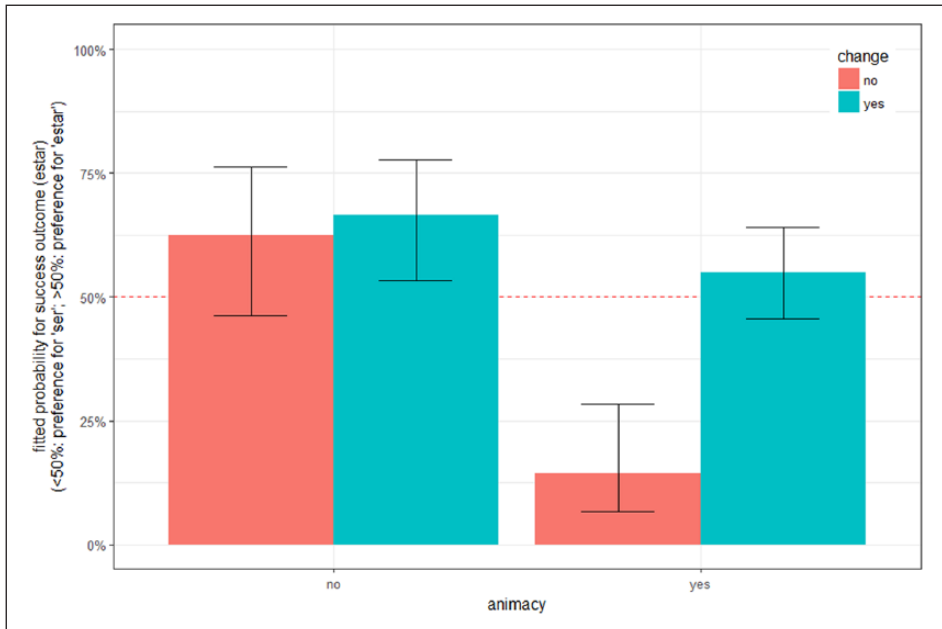


Figure 2. The selection of *estar* ‘to be’ with respect to the variable ‘susceptibility to change’ (no vs yes) and ‘animacy’ (no vs yes).

Figure 3 shows the interaction effect involving the ‘predicate type’ predictor and the ‘clause type’ predictor, a surprising effect since ‘clause type’ was only introduced because of its relevance in Romani. A cross-over pattern is observed: while with individual-level predicates in affirmative clauses respondents show a significant preference for *ser* and in negative clauses for *estar*, see (3a) and (3b) respectively, the opposite is true for stage-level predicates, that is, participants choose *estar* for affirmative clauses and mainly *ser* for negative clauses, see (3c) and (3d).

Individual predicate, affirmative clause

- (3a) Sí nuestro apartamento es grande.
 yes our apartment is(*ser*:3SG) big.3SG.M
 ‘Yes our apartment is big.’ (Item 6)

Individual predicate type, negative clause

- (3b) Sí, claro, mi papá no está estúpido.
 yes of_course my father NEG is(*estar*:3SG) stupid.3SG.M
 ‘Yes of course, my father isn’t stupid.’ (Item 10)

Stage predicate type, affirmative clause

- (3c) Porque no le llamé antes y ahora está enojado.
 because NEG 3SG.DAT call before and now is(*estar*:3SG) mad.3SG.M
 ‘Because I didn’t call him and now he is mad.’ (Item 2)

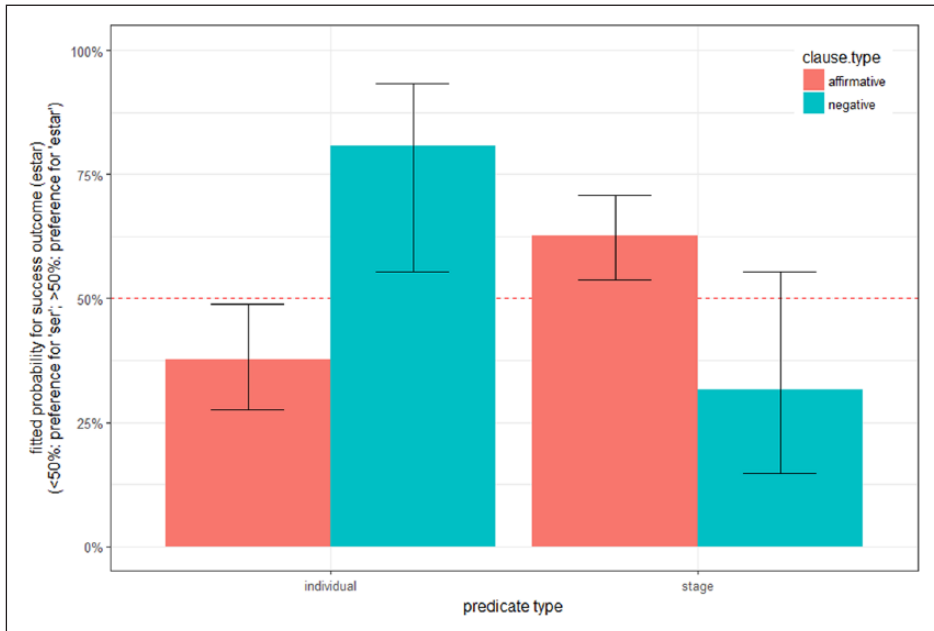


Figure 3. The selection of *estar* ‘to be’ with respect to the variable ‘predicate type’ (stage-levels vs individual-level) and ‘clause type’ (affirmative vs negative).

Stage predicate type, negative

- (3d) Sí, ahora tu coche no es muy grande.
 yes now your car NEG is(*ser*:3SG) very big.SG.M
 ‘Yes, now your car is not very big.’ (Item 25)

The next significant two-way interaction is between ‘predicate type’ and the language profile of the participant (see Figure 4). This interaction shows how the type of predicate can guide bilingual and monolingual speakers to make different copula choices. Neither group seems to significantly prefer either of the two variants when it comes to stage-level predicates, although there is a slight preference for *estar*, as in the clause *Sí, la cena está buena* ‘Yes, dinner is (*estar*) good’. The difference lies in the copula choice regarding individual-level predicates, for example, *No me gustó el dueño del apartamento, está/es desagradable* ‘I didn’t like the owner of the apartment, he is (*estar/ser*) unpleasant’. In these cases, monolinguals have a significant preference for *ser*, while bilinguals exhibit the same variable behaviour as with stage-level predicates.

In Figure 5, the interaction between ‘language’ and ‘susceptibility to change’ is plotted with respect to Spanish copula choice. It can be seen that when the trait is susceptible to change, as for example in *Cuando está/es alegre* ‘When he is (*estar/ser*) happy’, both bilinguals and monolinguals agree on using *estar* more often than *ser* (confidence intervals include the 50% threshold). On the contrary, if the adjective is not susceptible to change, as in *Ahora ella es/está católica también* ‘Now she is (*ser/estar*) Catholic too’, the behavior of the two groups differs. Although bilinguals and monolinguals both prefer *ser* over *estar*, they do so with different magnitudes (monolinguals: $100-20=80\%$ *ser*; bilinguals: $100-41=59\%$ *ser*), and while that preference is significant for the monolinguals, it is not for the bilinguals.

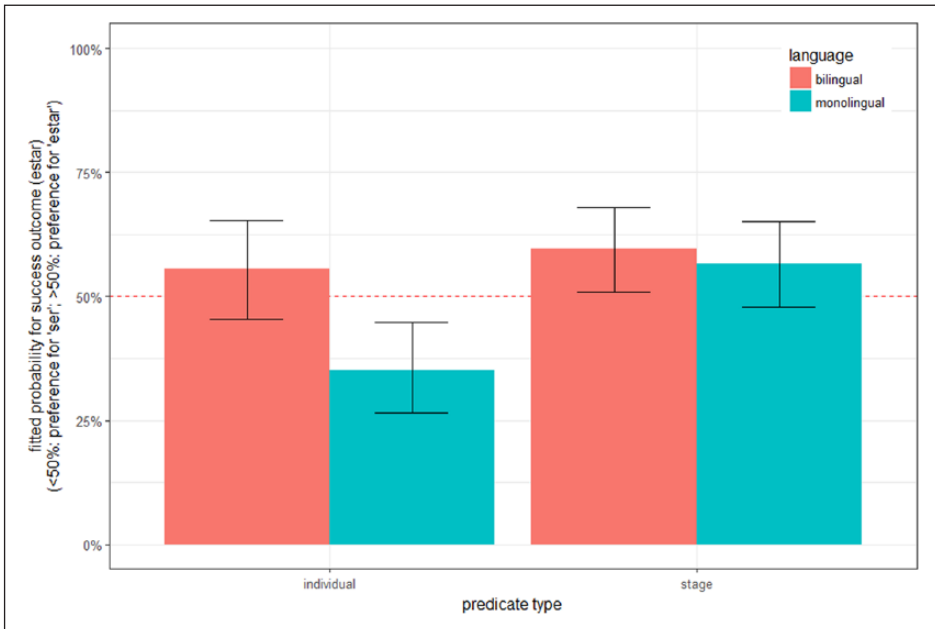


Figure 4. The selection of *estar* 'to be' with respect to the variable 'predicate type' (stage-level vs individual-level) and the language profile (bilingual vs monolingual).

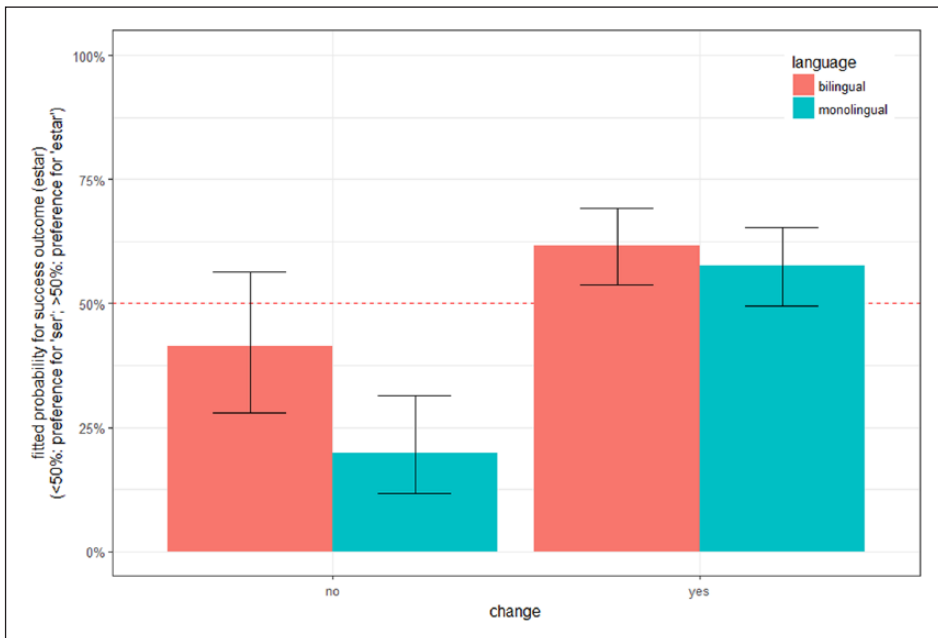


Figure 5. The selection of *estar* 'to be' with respect to the variable 'susceptibility to change' (no vs yes) and the language profile (bilingual vs monolingual).

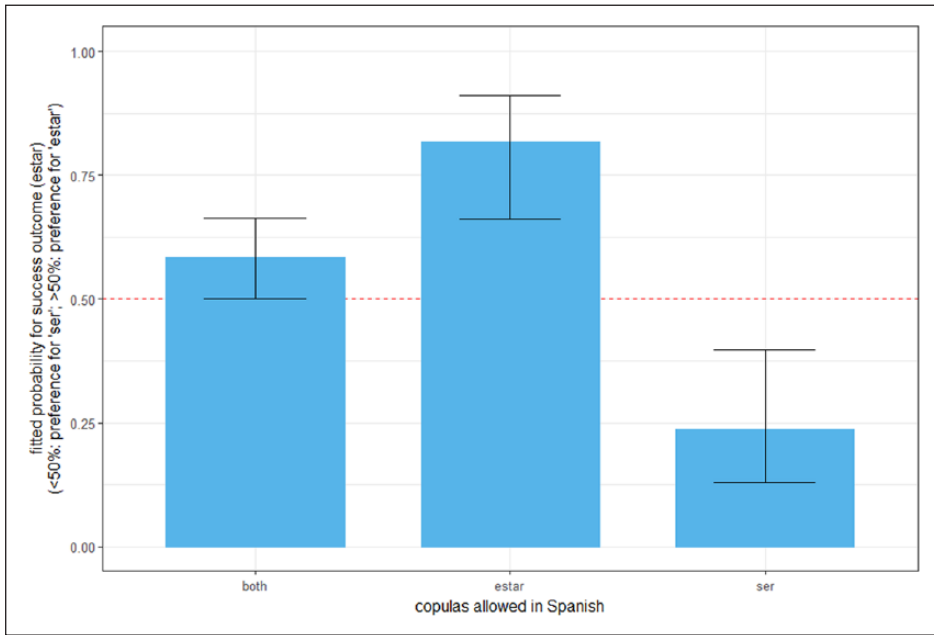


Figure 6. The selection of *estar* ‘to be’ with respect to the variable ‘copulas allowed in Spanish’ based on the adjective (both vs *estar* vs *ser*).

The last predictor that was included in the final model is ‘copulas allowed in Spanish’. In Figure 6, one can observe that, in line with expectations, participants overwhelmingly choose *ser* with adjectives that require *ser* and they significantly more often select *estar* when the adjective is used with *estar* among other Spanish-speaking groups. In addition, when ‘both’ copulas are possible, participants prefer to use *estar* in 58% of this type of clause. Given that the confidence interval for this estimate (i.e. 50%–66%) does not approach the 50% ‘no preference’ threshold, we could argue for a form of incipient generalization of *estar* when both copulas are allowed. No interaction between ‘copulas allowed in Spanish’ and ‘language’ could be retained, which highlights the fact that bilinguals are not more likely than monolinguals to generalize *estar* in clauses where adjectives can select both copulas.

The Romani and Spanish responses among the bilinguals and the Spanish responses among the monolinguals

In this sub-section, we aim to answer the study’s second research question: is simplification of alternatives in Spanish copula choice among bilinguals due to transfer from Romani? To address this question, we first consider the descriptive data for rates of selection. When taking into account only third person affirmative clauses, monolingual speakers alternate between the choice of *ser* and *estar*, with a slight preference for the latter (*ser* 45% vs *estar* 55%), whereas bilingual participants show a much more skewed preference for *estar* (*ser* 35% vs *estar* 65%). The Romani copula alternation is even more radical, since the *l-* clitics are chosen about 77% of the time as being the appropriate translation.

More evidence highlighting the incipient generalization of *estar* among the bilinguals and the potential role of the Romani *l-* clitics comes from the predictor ‘copulas allowed in Spanish’. This is the linguistic variable that codes the cases where some adjectives are more frequently encountered with the copula *ser* or *estar* or allow for more variation; see Figure 7 (clitics abbreviated as *lo* in the

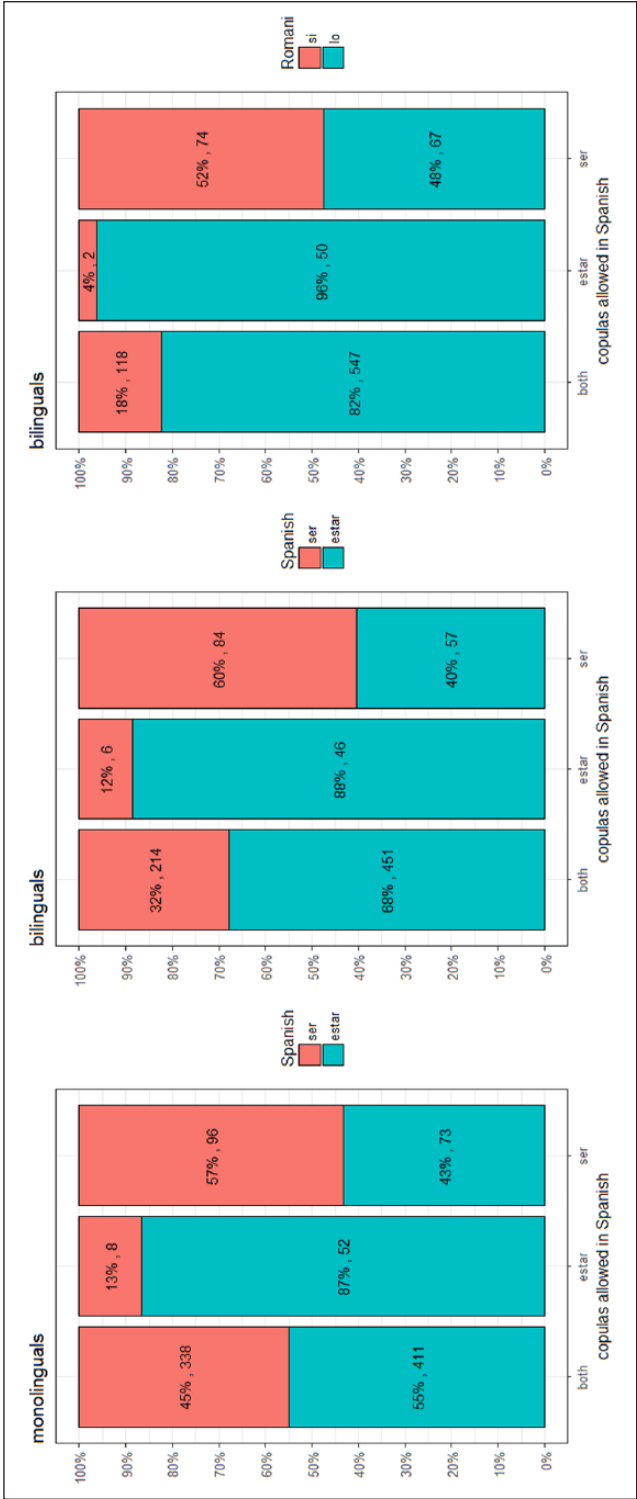


Figure 7. Copula choice for third person affirmative clauses among the monolinguals (in Spanish) and the bilinguals (in Spanish and in Romani) with respect to the variable ‘copulas allowed in Spanish’ depending on the adjective (both vs *estar* vs *ser*).

figures). It is unsurprising that participants overwhelmingly choose *estar* for adjectives where *estar* is the preferred option (see second column on the left and middle figures; 87% for the monolinguals and 88% for the bilinguals in Spanish). However, it is striking that they do not show this straightforward preference when it comes to adjectives in which *ser* is more frequently used among other Spanish-speaking groups. In this case, we note a more balanced picture (respectively 57% and 60% *ser* for the monolingual and bilingual speakers). Therefore, it appears that *estar* is gaining ground in both the monolingual and the bilingual groups from Mexico, at least in third person clauses.

In addition, for sentences where ‘both’ copulas are frequently used among other Spanish-speaking populations (first columns), we can see that we obtain a pattern in keeping with expectations as far as the monolinguals are concerned, that is, a near 50–50 split. An interesting picture emerges from the bilinguals’ responses, where *estar* is used in 68% of the ‘both’ cases and the *l*-clitics in 82% of those sentences. More generally, the distribution of the Romani copulas parallels that of their Spanish copula selection, with *l*-clitics chosen much more frequently for sentences with *estar* and in the vast majority of the cases in which both copulas would be appropriate.

One can therefore argue that the dynamism of the *l*-clitics in Romani third person affirmative clauses could account for the higher rates of selection of *estar* in the responses of the bilinguals as compared to the monolinguals. However, the question remains as to whether these similar frequency distributions are superficial similarities caused by different linguistic constraints, or whether the bilinguals’ underlying grammar of the Spanish copula alternation is indeed different from that of the monolinguals and more similar to the grammar that models the Romani copula variation. Only in the latter case should we be able to validate the role of conceptual transfer. Alternatively, it could be that the bilinguals’ generalization of *estar* is independent from copula choice in Romani, and that the bilinguals are closely following the constraints of the monolingual speakers. To address this question, we will examine the observed probabilities across the three datasets by relying on each of the significant interactions revealed by the generalized mixed-effects models for the responses of the monolinguals (see Appendix Table B.1).

In Figure 8, it can be seen that, as the *l*-clitics are generalized in the Romani data, the distinction between inanimate and animate referents is not important for individual-level predicates. Even though the Spanish responses of the Roma do not fully align with their Romani responses, the greater generalization of *estar* for individual-level predicates and inanimate referents among the Roma bilinguals when compared to the Spanish monolinguals could stem from the generalization of the Romani clitics in these contexts.

Example (4a) illustrates the preferred use of the Romani clitic with an inanimate referent and an individual-level predicate. Example (4b) illustrates the preferred use of the clitic with a stage-level predicate and an animate referent.

Individual predicate, inanimate

- (4a) o xabe lašo=**lo**
 DEF.SG.M food good-3SG.M
 ‘The dinner *is* good.’ (Item 5; SP 20 years old)

Stage predicate, animate

- (4b) akana xuljariko=**lo**
 now angry-3SG.M
 ‘Now he is mad.’ (Item 2; SP 18 years old)

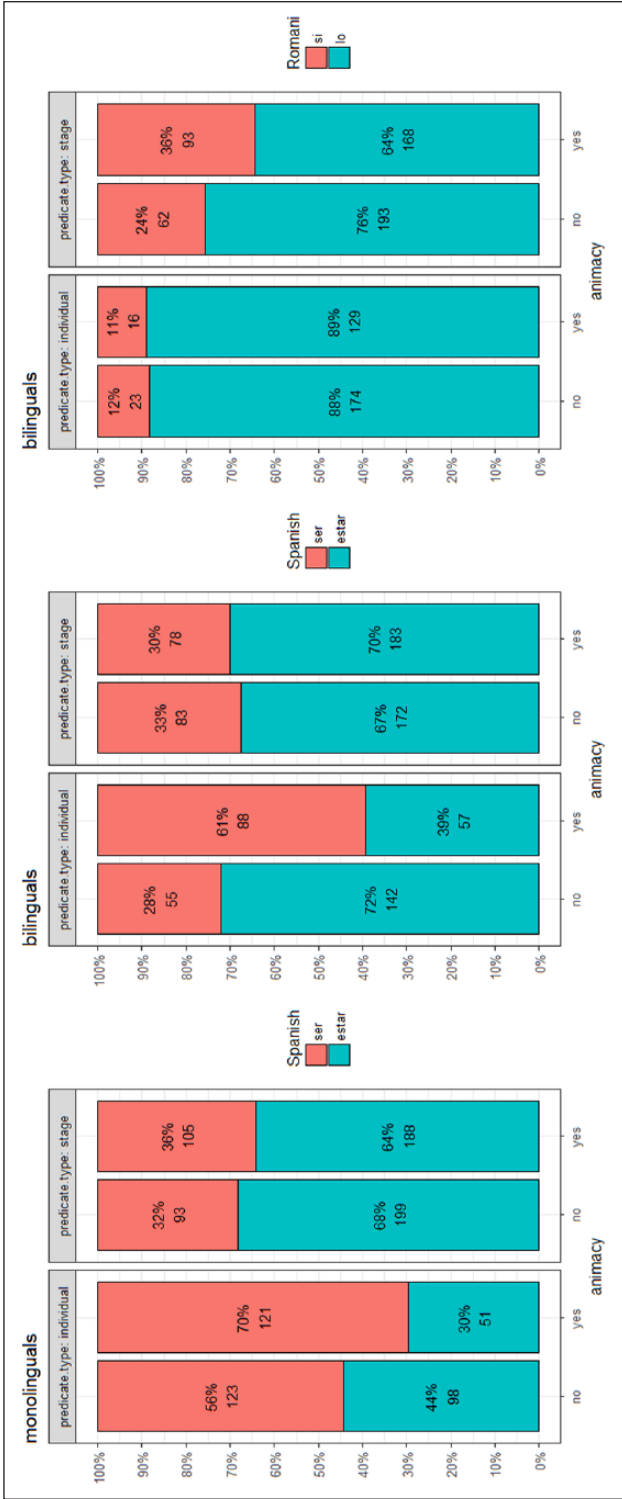


Figure 8. Barplots for the interaction 'predicate type x animacy'.

In the barplots in Figure 9, it can be seen that monolinguals and bilinguals behave in the same way with respect to Spanish, although it should be noted that the bilinguals prefer *estar* more often than the monolinguals, in particular for individual-level predicates. The more frequent selection of *estar* by the bilinguals could be the result of the more frequent selection of the *l-* clitics in Romani where it is almost categorical for individual-level predicates. In contrast, with stage-level predicates, the distinction between immediate and ongoing experience with the referent is much stronger in the Romani responses than in the Spanish responses of the bilinguals, which rather parallel the Spanish responses of the monolinguals.

Examples in (5) illustrate the preferred choices for stage-level predicates in Romani. Compare example (5a), where the speaker opts for the clitic in a clause that involves immediate experience with the referent combined with a stage-level predicate, and (5b) where the speaker opts for the copula and has ongoing, continuous experience with the person under discussion.

Stage-level predicate, immediate experience with the referent

- (5a) akana xuljariko=**lo**
 now angry-3SG.M
 ‘Now he is mad.’ (Item 2; SP 18 years old)

Stage-level predicate, ongoing experience with the referent

- (5b) akana bi voj **si** katolika
 now also 3SG.F.NOM be.3SG Catholic
 ‘Now she is Catholic too.’ (Item 21; SP 50 years old)

In Figure 10, it is shown that monolinguals prefer *estar* for stage-level predicates and *ser* for individual-level predicates, irrespective of the person number. This difference is levelled in the bilinguals’ Spanish responses and to some extent in their Romani responses. Again, the Romani responses for the individual-level predicates might help explain the higher rates of selection of *estar* in the bilinguals’ responses as compared to the monolinguals.

In Figure 11, one observes that overall the Spanish responses of the monolinguals and the bilinguals are similar, but that the bilinguals choose *estar* more frequently, in particular, for clauses involving ongoing experience with the referent and no change. The higher selection of *estar* by the bilinguals might be reinforced by the Romani responses (57% of the *l-* clitics for ongoing experience and no change).

In the barplots in Figure 12, it can be seen that the Spanish and the Romani responses of the bilinguals align very strongly and might account for the differences in the selection rates in their Spanish responses. More specifically, the bilinguals prefer the *l-* clitics in Romani when the experience with the referent is ongoing and the referent is inanimate (83%). This can explain the fact that they also choose *estar* in 70% of the cases in the same conditions, unlike monolinguals (only 34% for *estar*).

Finally, the barplots in Figure 13 show that the Spanish responses are similar among the monolinguals and the bilinguals. Again, the higher rates of *estar* for inanimates in the third person singular may be strengthened by the Romani predominance of the clitics in *l-* in such cases.

The last predictor examined is the relation between the semantic domains to which the predicate adjective belongs and the specific Romani or Spanish copula chosen by the participants (see Figures 14 and 15). It can be seen that for age and personality, the monolinguals and bilinguals behave alike in their Spanish copula selection with very few responses containing *estar*, but that the Romani responses significantly favour clitics. For colour, description, mental state, physical

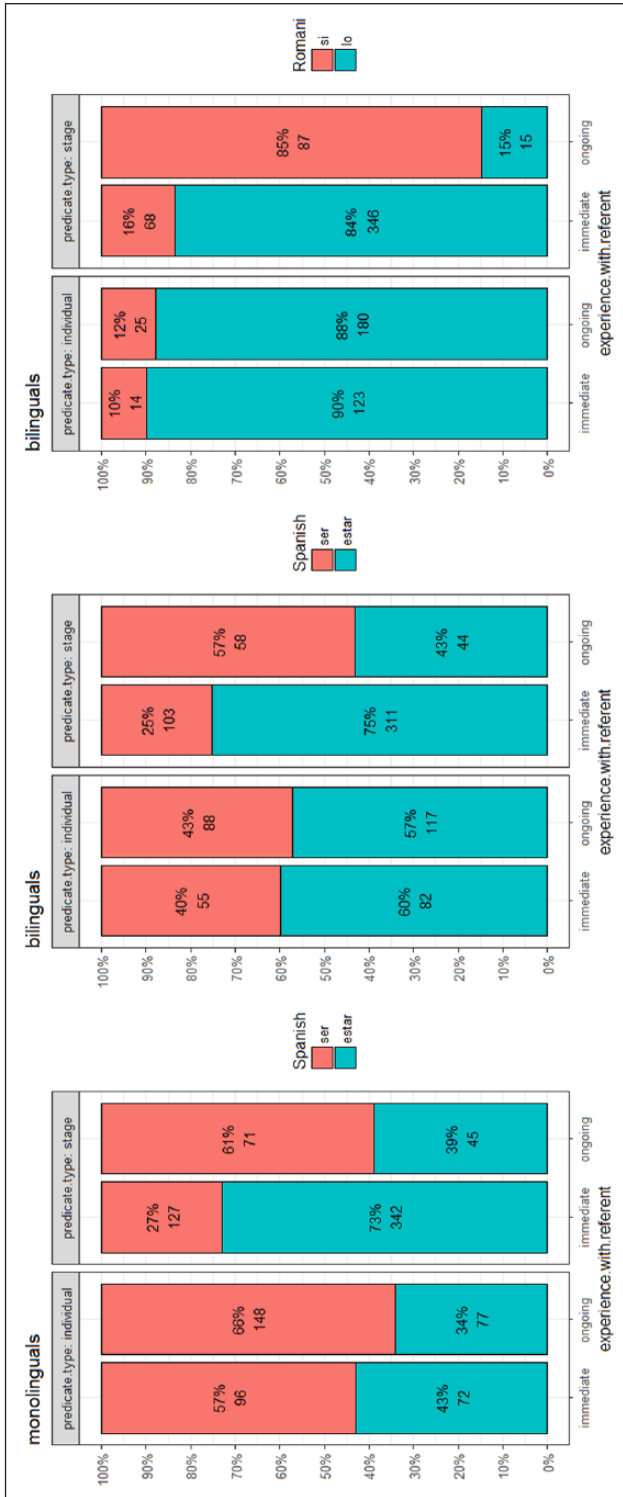


Figure 9. Barplots of interaction 'predicate.type x experience with referent'.

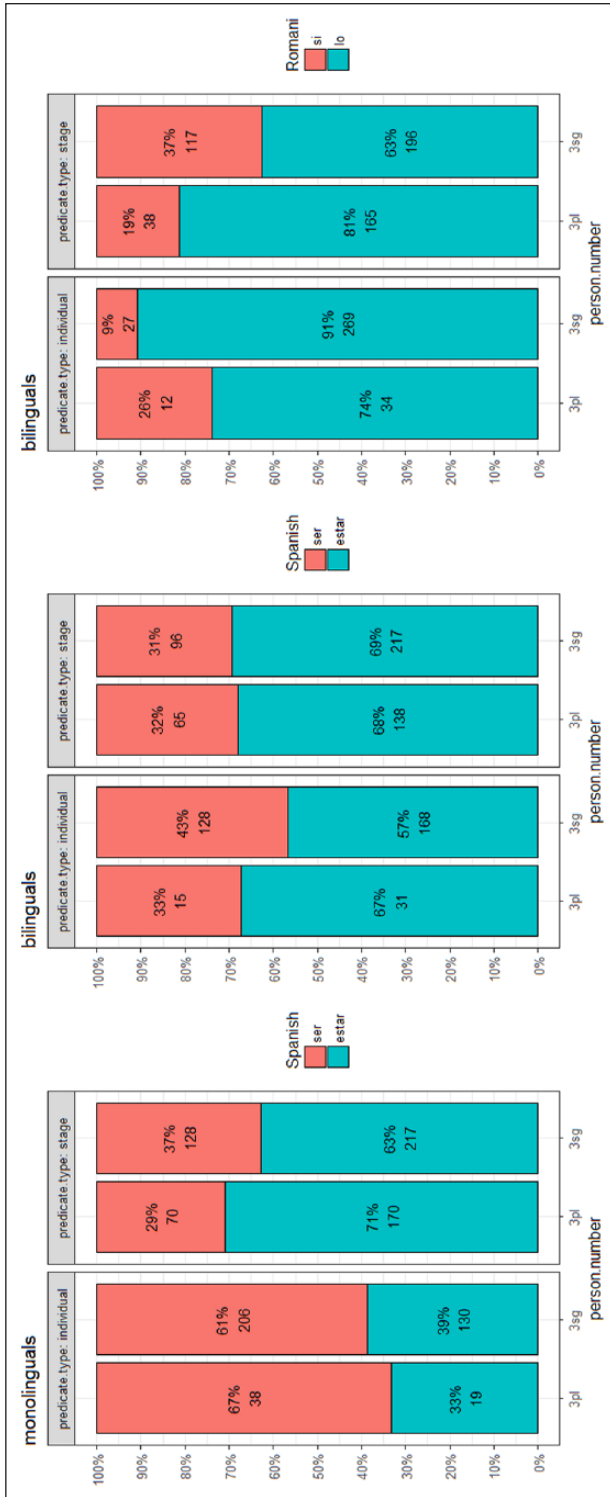


Figure 10. Barplots of the interaction 'predicate type x person number'.

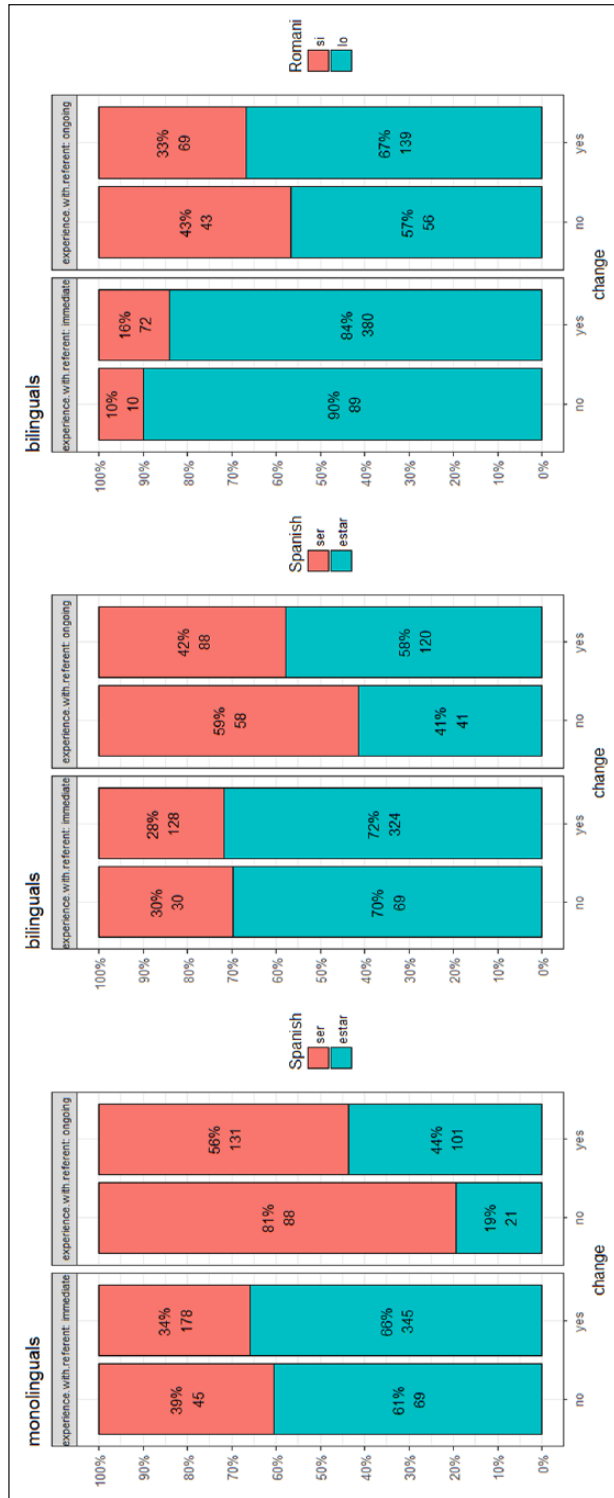


Figure 11. Barplots of the interaction ‘experience with referent x change’.

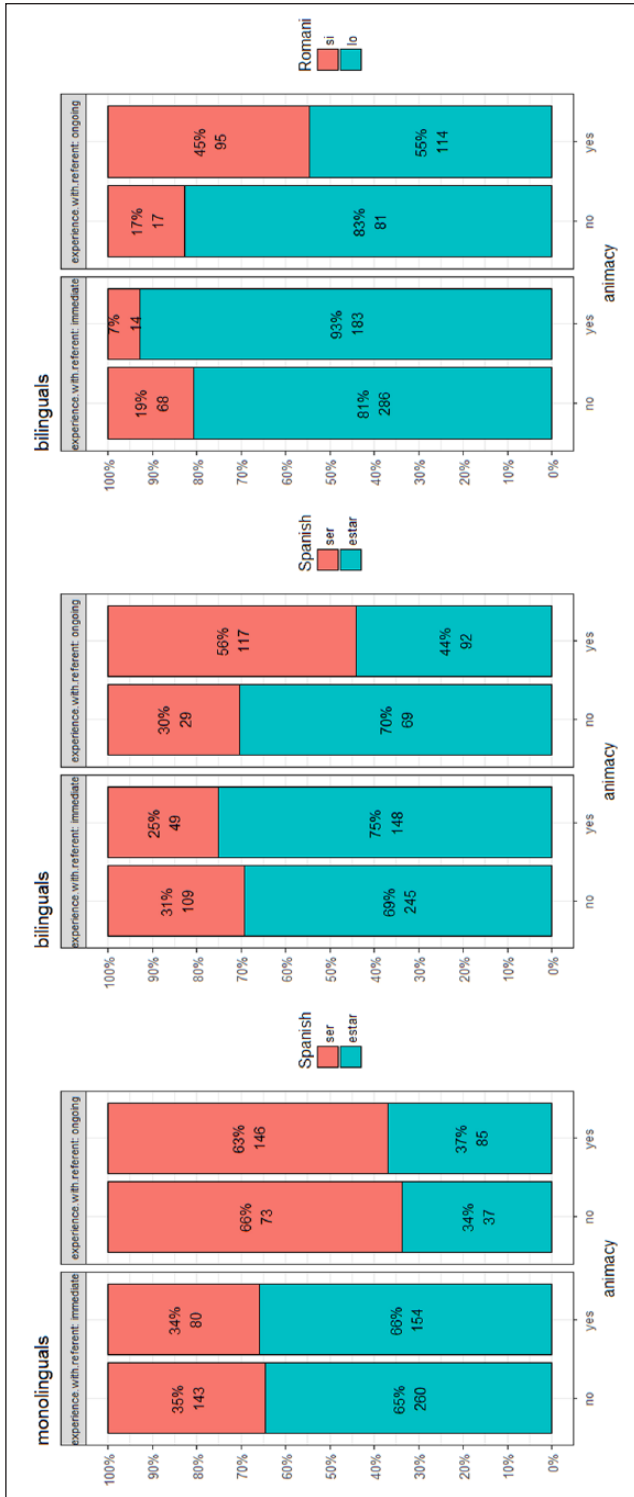


Figure 12. Barplots of the interaction 'experience with referent x animacy'.

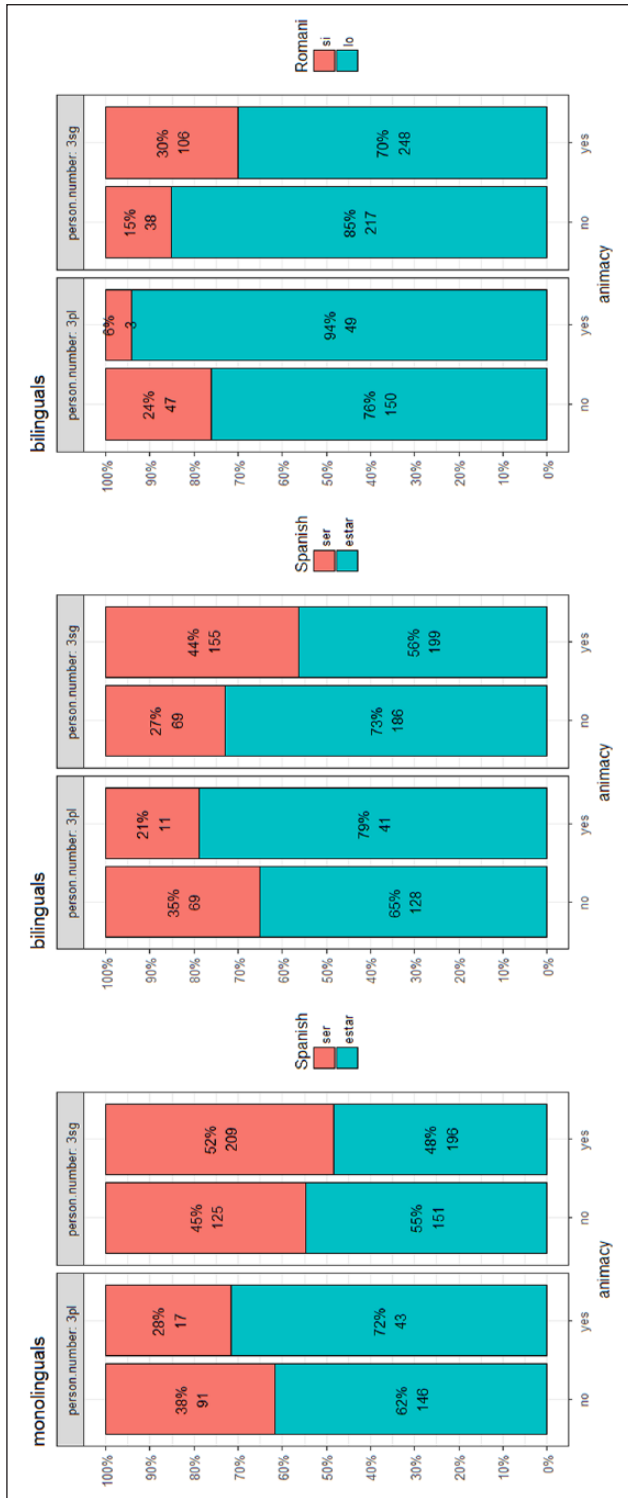


Figure 13. Barplots of the interaction 'person number × animacy'.

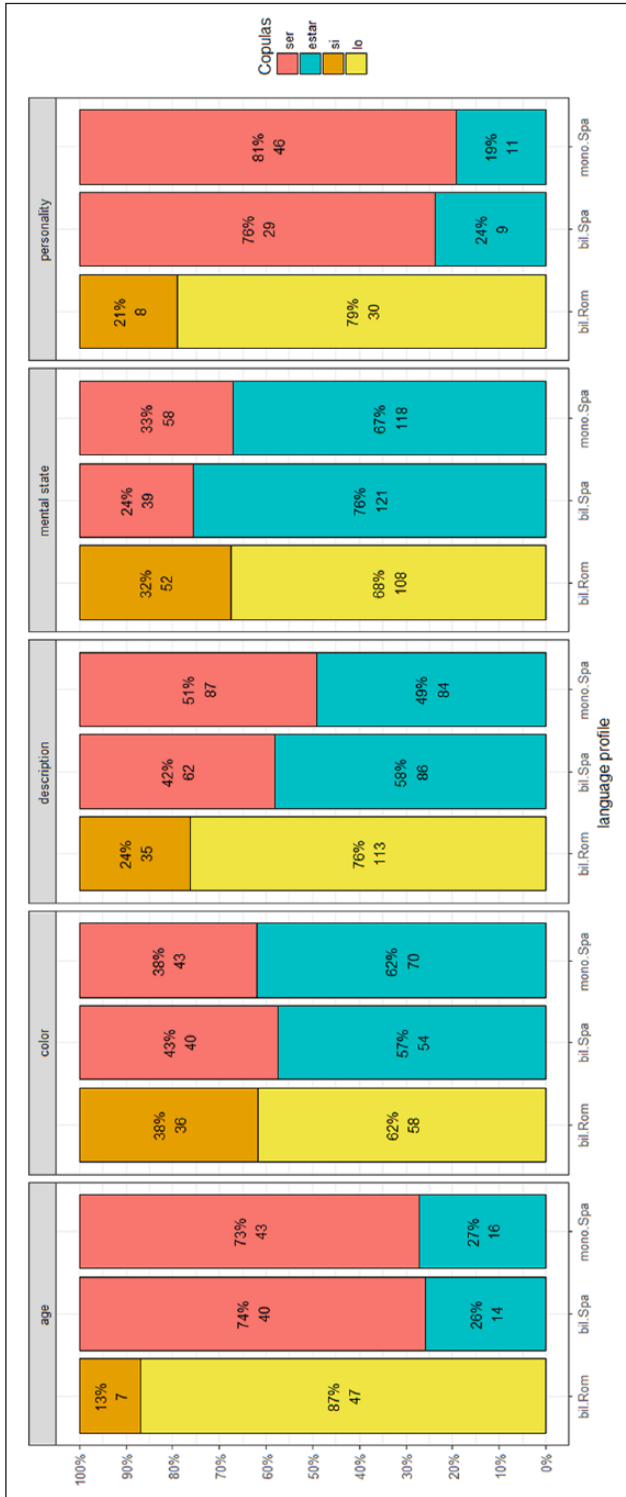


Figure 14. Barplot of the interaction between 'adjective class' and 'language profile', with response variable either the Romani copula alternation (in yellow and orange) or Spanish copula alternation (in blue and red). Adjective classes are age, colour, description, mental state, and personality.

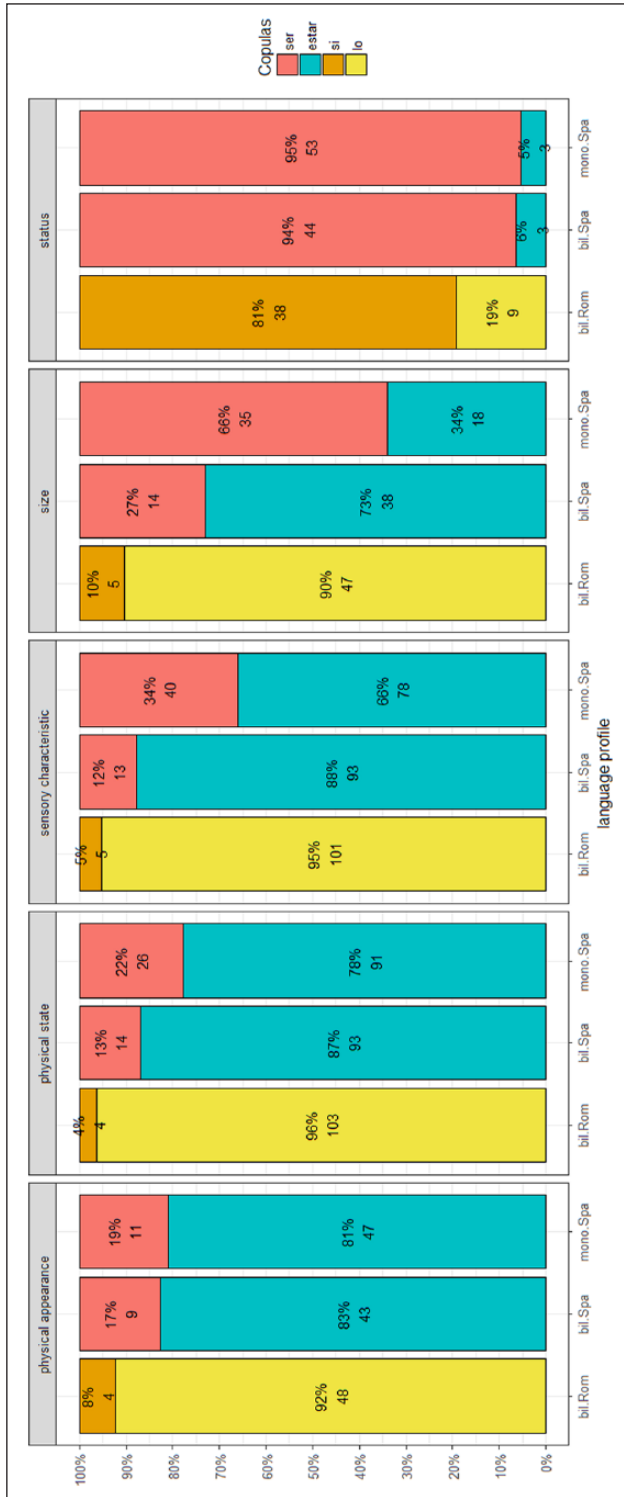


Figure 15. Barplot of the interaction between 'adjective class' and 'language profile', with response variable either the Romani copula alternation (in yellow and orange) or Spanish copula alternation (in blue and red). Adjective classes are physical appearance, physical state, sensory characteristic, size, and status.

appearance and physical state adjectives, the three sets of responses are very similar (Spanish from monolinguals and bilinguals and Romani responses). Finally, for sensory characteristics and size, the Spanish monolingual group behaves differently by selecting *estar* less frequently than bilinguals do in both Romani and Spanish.

Discussion

This paper addresses two research questions. The first question is whether there is simplification of alternatives among bilinguals as compared to monolinguals. In particular, we test whether the same simplification processes reported in Silva-Corvalán (1986) are also evidenced in the Spanish copula choice among the Romani–Spanish bilinguals in Mexico, or whether no such tendency is observed as in some of the bilingual communities examined in Geeslin and Guijarro-Fuentes (2008). The second research question is whether unidirectional or bidirectional transfer between the two languages might account for these results.

First, we have examined how different linguistic variables and their joint action affect the choice of either *estar* ‘to be’ or *ser* ‘to be’. In particular, we were interested in uncovering how the influence of these predictors could vary between monolinguals and bilinguals, revealing possible differences in the underlying grammar of copula variation. Analysis of the results shows that both groups frequently select *estar* in contexts where *ser* could be expected and in contexts where both copulas are possible. In addition, for both groups ‘susceptibility to change’ interacting with ‘animacy’ as well as ‘predicate type’¹⁰ interacting with ‘clause type’ are relevant combinations of variables for the choice of the Spanish copulas.

When comparing the Mexican results to those of the Spanish speakers from Spain (Geeslin & Guijarro-Fuentes, 2008), we note that only ‘predicate type’ is relevant for all groups in both studies. However, in our study ‘predicate type’ interacts with ‘clause type’, a novel variable that we introduced in our research because of Romani and that had not been previously discussed in the literature on Spanish. Geeslin and Guijarro-Fuentes (2008) additionally find that ‘copulas allowed’ and ‘adjective class’ are relevant factors for all the groups under study. Although ‘copulas allowed’ is also relevant for all Mexican Spanish speakers in our study, we do not have any comparable results on ‘adjective class’ as we did not include this variable in our statistical analysis. The factor ‘susceptibility to change’ in interaction with ‘animacy’ is significant for both groups in the current study. ‘Susceptibility to change’ alone was significant for the Catalans, Valencians and monolinguals in the study by Geeslin and Guijarro-Fuentes (2008) and ‘animacy’ was only significant for the monolingual group. An interesting insight comes from Geeslin (1999), who notes that the variables ‘adjective class’ and ‘animacy’ largely overlap; for example, the adjective classes that favour *estar* correspond to animate referents. Because of this overlap, only one variable is significant in a given statistical model. The fact that we did not include adjective class in our model may have allowed the variable ‘animacy’ to gain significance. Finally, ‘experience with the referent’, which was not significant in our study, was found to be significant for the Catalans, Galicians, and monolinguals from the Iberian Peninsula.¹¹

With respect to the first research question, that is, whether bilinguals generalize the use of the copula *estar* more than their monolingual counterparts, the overall rates suggest that such generalization is indeed taking place (bilinguals choose *estar* in 55.77% of responses vs monolinguals 45.85%). However, the statistical analysis does not offer any support to the generalization of *estar* by the bilinguals. Selection rates of *estar* among the bilinguals in our study are higher than the rates of the bilingual groups from Spain, that is, Basques (47.3%), Catalans (43.8%), Galicians (48.7%) and Valencians (44.5%). In contrast, selection rates of *estar* among the monolinguals in our study are close to those of the monolinguals from Spain (44.9%) (Geeslin & Guijarro-Fuentes, 2008: 371). Differences in *estar* selection rates between our bilinguals and the other bilingual groups

from Spain might be due to geographical differences or differences in education levels, as our bilingual participants had significantly lower education levels than those from Spain. Interestingly, the monolingual respondents in our study resemble the monolingual respondents from Spain despite differences in education levels (secondary vs higher education respectively).

To evaluate the extent of generalization of *estar*, however, one needs to also examine the precise linguistic conditioning, not mere rates. A closer look at the data shows that bilinguals behave differently from monolinguals as far as the variables ‘predicate type’ (as well as ‘frame of reference’) and ‘susceptibility to change’ are concerned by generalizing *estar* in contexts where monolinguals prefer *ser*. This is the case, in particular, for individual-level predicates, when the attribute applies to the referent as a whole (or with a class frame of reference), and when the relationship between the referent and the attribute is not susceptible to change. Interestingly, Silva-Corvalán (1986) identified both frame of reference and susceptibility to change as factors that drive the simplification process among the heritage speakers of Spanish in Los Angeles. We can therefore conclude that Romani bilinguals extend the use of *estar* in contexts where Spanish monolinguals still prefer the copula *ser*.

Regarding the second research question, that is, whether extension of *estar* is due to transfer, we worked with a dataset restricted to third person affirmative clauses where variation exists in Romani. We noted that clitics are the preferred option in Romani third person affirmative clauses, with 77% of use in translation. We observed, in parallel, that the Romani bilinguals frequently prefer *estar* in these clauses (65% as opposed to 55% for the monolinguals), suggesting that transfer from Romani may be playing a role in copula choice in Spanish. In addition, it is clear that we cannot account for these results in terms of priming from Spanish *estar* to the Romani clitics, as the Romani responses using the *l-* clitics outnumber the mere translation of the Spanish clauses with *estar*.

When examining the linguistic conditioning of the variation, analysis of our results shows that, overall, in third person clauses Romani bilinguals rely on the same variables in Spanish as the monolingual speakers do. A careful comparison of the rates of selection of *estar* among the monolinguals and the bilinguals shows that the bilinguals opt for *estar* more frequently in some contexts, for example, when the experience with the referent is ongoing and the referent is inanimate, as well as for size and sensory characteristic adjectives. The high rates of use of the Romani clitics in those same contexts could account for the expansion of *estar* in most of these cases. It is therefore possible to consider the Romani clitics as triggers of the generalization of *estar*, or at least catalyzers. To confirm this claim, a new study is needed, designed to test more specifically the Romani contexts of variation and the Spanish equivalents.

Conclusion

Adamou (2013) highlighted the long-term effects of partial conceptual equivalence encoded by distinct linguistic means that led to the transfer of Spanish copula variation in the heritage Romani language. In Adamou (2013), however, the focus was on a contact-induced outcome in a historical perspective indicating the tendency for two languages to converge through conceptual transfer. The present study looks more closely at the level of the individual speaker and shows that, even though the two grammars of the Romani–Spanish bilinguals do not fully align, bidirectional conceptual transfer seems to be ongoing. Taken together, the evidence from these studies suggest that online, cognitive pressures at the level of the bilingual to reduce the cognitive load may lead to contact-induced language change in the long run as the two languages converge. This is particularly true in small communities with little or no normative pressure and no exposure to competing monolingual input as is the case for the Romani speakers in Mexico, unlike what is generally observed among the Spanish–English bilinguals in the USA. This convergence process through transfer may lead to the complexification of one system, as is the case for Romani grammar, which has added variation at copula level, or simplification at some second stage, by reducing the variation in both Romani and

Spanish. To account for these conflicting processes, we suggest that, as the two grammars become similar, reduction of cognitive load is overall achieved at the level of the bilingual speaker. In the future, online processing studies might help us shed more light on the cognitive mechanisms at play and test the reduced cognitive load hypothesis in cases of similar choices across the two languages.

Acknowledgements

We thank all the participants in the study. In particular, we thank members of the Romani community of La Rinconada for their hospitality. In Mexico City, we thank Carmina Icaza Conde, Josefa, and Manuel de los Reyes García Martínez (RIP). We also thank the reviewer for the careful reading of the paper and for insightful comments. Preliminary versions of this paper were presented by one or several authors at Bilingualism in the Hispanic and Lusophone World, International Conference on Romani Linguistics, International Conference on Language Variation in Europe, and the Department of Linguistics at the University of Boulder, Colorado. We are grateful to members of those audiences for useful suggestions. All errors are our responsibility.

Declaration of conflicting interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Data collection was partly supported by the French National Centre for Scientific Research (CNRS) and the INALCO, via the laboratory Oral Tradition Languages and Civilizations (LACITO UMR 7107) to Cristian Padure. This research was also supported by the excellence cluster Empirical Foundations of Linguistics (Labex EFL), funded by the French National Research Agency (Investments for the Future, anr-10-labx-0083), to permanent member Evangelia Adamou and visiting doctoral fellow Stefano De Pascale.

Notes

1. The lower estimates are reported in the *Revue des Etudes Tsiganes* (2012) and the higher estimates are provided by the Romani organization SKOKRA.
2. 1, 2, 3: first, second, third person; DAT: dative; DEF: definite; F: feminine; LOC: locative; M: masculine; NEG: negative; NOM: nominative; PL: plural; POSS: possessive; SG: singular.
3. ‘Circumstantiality’, ‘semantic transparency’ and ‘adjective class’ are also good predictors, but we did not investigate them in the current study.
4. A reviewer observes that a monolingual control group from Veracruz would have been more appropriate as there may be differences in the varieties of local Spanish. Although we agree that this would have been ideal, to our knowledge, there is no evidence in the literature reporting differences in the uses of the copulas in various areas of Mexico.
5. We do not expect this difference to have a significant effect on copula choice as both studies by Gutiérrez (1994) and Cortés-Torres (2004) showed that speakers with primary and secondary education are more similar to one another but differ greatly from speakers with college education. More specifically, Gutiérrez (1994) reports 28% of *estar* uses for speakers with primary education versus 25% for those with secondary education. Cortés-Torres (2004) found similar results: speakers with no education or primary education favoured *estar* in 30% of cases, those with technological education in 28% of cases, followed by those with secondary education in 25% of cases, contrasting with speakers with college education (or *preparatoria* level) who used *estar* in only 16% of cases.
6. In terms of raw numbers (not combinations of variables) we worked in the reduced version with 7 items for individual predicates versus 10 for stage predicates (initial version: 17 individual vs 12 stage); 11 for immediate experience versus 6 for ongoing experience (initial version: 17 immediate vs 12 ongoing); 4 items for no change versus 13 for change (initial version: 8 for no change vs 21 change); 9 items for inanimates versus 8 for animates (initial version: 11 inanimates vs 18 animates); all 10 adjective classes were represented by at least one item.

7. For the sake of consistency, all models were fitted with the optimizer ‘bobyqa’ (default: Nelder-Mead) and by setting the number of iterations at 50,000 (default: 10,000).
8. Frame of reference and predicate type are two variables that completely overlap and therefore only predicate type was maintained in the model.
9. In order to avoid data sparseness issues and to create robust models, we only examined predictor interaction when there were sufficient data points for every level combination (minimum cell frequency of 15). Therefore, we ruled out all possible three-way interactions with ‘copula choice allowed’, only selecting its two-way interaction with ‘language’.
10. With the equivalent ‘frame of reference’, a variable excluded from the model as it overlapped with ‘predicate type’.
11. Geeslin and Guijarro-Fuentes (2008: 373) note that ‘experience with the referent’ is only significant with a large number of tokens so this might explain the lack of significance in our study.

References

- Acuña, E., & Adamou, E. (2013). *How did linguistic innovations spread in the Romani varieties of the Americas? A social interconnection approach*. Paper presented at the International Conference on Romani Studies, 12–14 September 2013, University of Strathclyde, Glasgow.
- Adamou, E. (2013). Replicating Spanish *estar* in Mexican Romani. *Linguistics*, 51(6), 1075–1105.
- Adamou, E., & Shen, R. X. (2017). There are no language switching costs when codeswitching is frequent. *International Journal of Bilingualism*. Epub ahead of print 14 May 2017. doi: 10.1177/1367006917709094.
- Arche, M. (2006). *Individuals in time*. Amsterdam: John Benjamins.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.
- Brown, E., & Cortés-Torres, M. (2012). Syntactic and pragmatic usage of the [*estar* + Adjective] construction in Puerto Rican Spanish: ¡Está brutal! In K. Geeslin & M. Díaz-Campos (Eds.), *Selected proceedings of the 14th hispanic linguistics symposium* (pp. 61–74). Somerville, MA: Cascadilla Proceedings Project.
- Bylund, E. (2009). Maturational constraints and first language attrition. *Language Learning*, 59, 687–715.
- Camacho, J. (1997). La distinción aspectual entre *ser* y *estar*. In A. Mendikoetxea & M. Uribe-Etxebarria (Eds.), *Theoretical issues at the morphology-syntax interface* (pp. 93–100). Bizkaia, Spain: Universidad del País Vasco.
- Camacho, J. (2015). What do Spanish copulas have in common with Tibetan evidentials? In S. Gumiel-Molina, M. Leonetti, & I. Pérez-Jiménez (Eds.), *New perspectives on the study of Ser and Estar* (pp. 173–202). Amsterdam: John Benjamins.
- Clements, C. J. (1988). The semantics and pragmatics of the Spanish <copula + adjective> construction. *Linguistics*, 26, 779–822.
- Cortés-Torres, M. (2004). ¿*Ser* o *estar*? La variación lingüística y social de *estar* más adjetivo en el español de Cuernavaca, México. *Hispania*, 87(4), 788–795.
- De Jonge, R. (1993). (Dis)continuity in language change: *Ser* and *estar* + age in Latin American Spanish. In: F. Drikkonigen & K. Hengeveld (Eds.), *Linguistics in the Netherlands* (pp. 69–80). Amsterdam, the Netherlands: John Benjamins.
- Delbecque, N. (2000). Las cópulas *ser* y *estar*. Categorización frente a deixis. *Revista española de lingüística aplicada*. Extra 1, 239–280.
- Díaz-Campos, M., & Geeslin, K. L. (2011). Copula use in the Spanish of Venezuela: Social and linguistic sources of variation. *Spanish in Context*, 8(1), 73–94.
- Elšik, V., & Matras, Y. (2006). *Markedness and language change*. New York, NY: Mouton de Gruyter.
- Escandell-Vidal, V. (2017). Evidential commitment and feature mismatch in Spanish *estar* constructions. *Journal of Pragmatics*, 128, 102–115.
- Escandell-Vidal, V., & Leonetti, M. (2002). Coercion and the stage/individual distinction. In J. Gutiérrez-Rexach (Ed.), *From words to discourse* (pp. 159–179). New York, NY: Elsevier.
- Etudes Tsiganes*. (2012). Special issue ‘Les tsiganes en Amérique du Sud’, 51.

- Falk, J. (1979). *Ser y estar con atributos adjetivales*. Uppsala: Almqvist & Wiksell.
- Fernández, L., & María, J. (1995). Las construcciones con el verbo *estar*: aspectos sintácticos y semánticos. *Verba*, 22, 253–284.
- García-Márkina, Y. (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*. PhD thesis at the University Sorbonne Nouvelle-Paris 3.
- Geeslin, K. L. (1999). *The second language acquisition of copula choice in Spanish and its relationship to language change* (PhD thesis). The University of Arizona.
- Geeslin, K. L. (2001). Changing norms, moving targets and the SLA of copula choice. *Spanish Applied Linguistics*, 5(1–2), 29–55.
- Geeslin, K. L., & Guijarro-Fuentes, P. (2008). Variation in contemporary Spanish: Linguistic predictors of 'estar' in four cases of language contact. *Bilingualism: Language and Cognition*, 11(3), 365–380.
- Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, 58, 787–814.
- Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, 25, 515–530.
- Grosjean, F. (2011). An attempt to isolate, and then differentiate, transfer and interference. *International Journal of Bilingualism*, 16, 11–21.
- Guijarro-Fuentes, P., & Geeslin, K. L. (2006). Copula choice in the Spanish of Galicia: The effects of bilingualism on language use. *Spanish in Context*, 3(1), 63–83.
- Gutiérrez, M. J. (1994). 'Ser' y 'estar' en el habla de Michoacán, México. Mexico: UNAM.
- Jarvis, S., & Pavlenko, A. (2008). *Cross-linguistic influence in language and cognition*. New York, NY: Routledge.
- Kroll, J., & Dussias, P. (2013). The comprehension of words and sentences in two languages. In K. Bhatia Tej & W. C. Ritchie (Eds.), *The handbook of bilingualism and multilingualism* (pp. 216–243). Malden, MA: Wiley-Blackwell.
- Kroll, J., Dussias, P., Bice, K., & Perroti, L. (2015). Bilingualism, mind, and brain. *Annual Review of Linguistics*, 1, 377–394.
- Lema, J. (1992). Distinguishing copular and aspectual auxiliaries: Spanish *Ser* and *Estar*. In J. Amastae, J. Goddall, M. Montalbetti, & M. Phinney (Eds.), *Contemporary research in Romance Linguistics* (pp. 257–274). Amsterdam, the Netherlands: John Benjamins.
- Luján, M. (1981). The Spanish copulas as aspectual indicators. *Lingua*, 54, 165–210.
- Maienborn, C. (2005). A discourse based account of Spanish *ser/estar*. *Linguistics*, 43(1), 155–180.
- Malaver, I. (2012). Estamos muy grandes ya. Adjetivos de edad con 'ser' y 'estar' en el español de México y Guatemala. *Lexis*, 36(2), 191–224.
- Marín Gálvez, R. (2004). *Entre ser y estar*. Madrid, Spain: Arco-Libros.
- Matras, Y. (2000). The structural and functional composition of Romani demonstratives. In V. Elšik & Y. Matras (Eds.), *Grammatical relations in Romani: The noun phrase* (pp. 95–122). Amsterdam, the Netherlands: John Benjamins.
- Matras, Y. (2002). *Romani: A linguistic introduction*. Cambridge, UK: Cambridge University Press.
- Matras, Y. (2005). The classification of Romani dialects. A geographic-historical perspective. In D. Halwachs & B. Schrammel (Eds.), *General and applied Romani linguistics* (pp. 7–26). Munich, Germany: Lincom Europa.
- Montrul, S. (2006). On the bilingual competence of Spanish heritage speakers: Syntax, lexical-semantics and processing. *International Journal of Bilingualism*, 10, 37–69.
- Navas Ruiz, R. (1963). *Ser y estar: estudio sobre el sistema atributivo del español*. Salamanca, Spain: Acta Salmanticensia.
- Ortiz López, L. (2000). Extensión de 'estar' en contextos de 'ser' en el español de Puerto Rico. *Boletín de la academia puertorriqueña de la lengua española* (pp. 98–118).
- Padure, C., De Pascale, S., & Adamou, E. (in press). Variation between the copula *si* 'to be' and the *l*-clitics in Romani spoken in Mexico. *Romani Studies*.
- Porroche Ballesteros, M. (1988). *Ser, estar y verbos de cambio*. Madrid, Spain: Arco Libros.

- R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Schmid, M. S., & Köpcke, B. (2017). The relevance of first language attrition to theories of bilingual development. Epistemological keynote paper with commentaries. *Linguistic Approaches to Bilingualism*, 7(6): 637–667.
- Schmitt, C. (1992). *Ser and estar*: A matter of aspect. In *NELS 22 Proceedings* (pp. 411–426). Amherst, MA: GLSA.
- Silva-Corvalán, C. (1986). Bilingualism and language change: The extension of *estar* in Los Angeles Spanish. *Language*, 62, 587–608.
- Silva-Corvalán, C. (1994). *Language contact and change: Spanish in Los Angeles*. Oxford, UK: Clarendon Press.
- Soares, C., & Grosjean, F. (1984). Bilinguals in a monolingual and a bilingual speech mode: The effect on lexical access. *Memory and Cognition*, 12, 380–386.
- Sorace, A., & Serratrice, L. (2009). Internal and external interfaces in bilingual language development: Beyond structural overlap. *International Journal of Bilingualism*, 13, 195–210.
- Torres Cacoulios, R., & Travis, C. (2018). *Bilingualism in the community. Code-switching and grammars in contact*. Cambridge, UK: Cambridge University Press.
- Vaño Cerdá, A. (1982). *Ser y estar + adjetivos. Un estudio diacrónico y sincrónico*. Tübingen, Germany: Gunter Narr Verlag.

Author biographies

Evangelia Adamou is senior researcher at the National Centre for Scientific Research, France, in the LACITO research unit. She specializes in the analysis of lesser-known languages with a focus on language contact, combining corpus and experimental methods. She is the author of more than 60 scholarly publications including a monograph entitled *A corpus-driven approach to language contact: Endangered languages in a comparative perspective* (2016, Mouton de Gruyter). Her work has been supported by grants from the French National Research Agency.

Stefano De Pascale started working on his PhD project in 2015 at the QVLV Research Unit with a KU Leuven scholarship, under the supervision of Stefania Marzo and Dirk Speelman. In his research he investigates which contribution token-based semantic vector space models can make to the lexicometric study of lexical variation. Broadly speaking, his research interests involve the quantitative modeling of linguistic and sociocultural variation.

Yekaterina García Márkina is assistant professor of Spanish Linguistics at the University of Tours, France. She received her PhD in Linguistics from the University Sorbonne-Nouvelle Paris 3. Her research centers on variation in Spanish, and more specifically the use of *ser* and *estar* with adjectival predicates. She is also keenly interested in contrastive linguistics (Spanish-French).

Cristian Padure is preparing his PhD at the INALCO, in the LACITO research unit, under the supervision of Evangelia Adamou. In his thesis, he investigates copula choice among Roma in the community of La Rinconada, Mexico. He has taught Romani, his native language, at the INALCO and various NGOs.

Appendix A

Table A.1. Model estimates for the generalized linear mixed-effects regression on the full dataset (all persons and all sentences) and Spanish copula choice as response variable.

Term	Estimate	STD Error	Statistic (z-value)	p value
(Intercept)	0.8211	0.3869	2.122	0.03382 *
Clause type	1.9389	0.7227	2.683	0.00730 **
Affirmative – negative				
Predicate type	0.6515	0.3236	2.0313	0.04408 *
Individual – stage				

(Continued)

Table A.I. (Continued)

Term	Estimate	STD Error	Statistic (z-value)	p value
Change	-0.2796	0.4523	-0.618	0.53650
No change – change				
Animacy	-2.2967	0.5578	-4.117	3.83e-05 ***
Inanimate – animate				
Language	-1.4639	0.2004	-7.307	2.74e-13 ***
Bilingual – monolingual				
Copulas allowed in spanish	1.1572	0.4466	2.591	0.00957 **
Both – estar				
Both – ser	-1.5046	0.4603	-3.269	0.00108 **
Predicate type : clause type	-3.2290	1.0064	-3.208	0.00133 **
Individual : affirmative – stage : negative				
Change : animacy	1.8087	0.6549	2.762	0.00575 **
No change : inanimate – change : animate				
Change : language	0.8793	0.1925	4.569	4.90e-06 ***
No change : bilingual – change : monolingual				
Predicate type : language	0.7125	0.1675	4.254	2.10e-05 ***
Individual : bilingual – stage : monolingual				

* $p < .05$; ** $p < .01$; *** $p < .001$.

Appendix B

Table B.1 shows the output of the three different regression models. Starting with the same linguistic predictors for building the three generalized linear mixed-effects models: that is ‘predicate type’ (stage vs individual), ‘experience with referent’ (immediate vs ongoing), ‘susceptibility to change’ (no vs yes), ‘animacy’ (no vs yes), we added ‘person number’ (third singular vs third plural).

Table B.I. The significance of linguistic predictors across models.

	Monolinguals (Spanish)	Bilinguals (Spanish)	Bilinguals (Romani)
Experience with referent	***	NS	*
Predicate type	*	NS	*
Predicate type × animacy	***	**	NS
Predicate type × experience with referent	**	*	NA
Predicate type × person number	***	NS	NA
Experience with referent × change	***	NS	NA
Experience with referent × animacy	*	NS	NA
Person number × animacy	*	NA	NA
Variance explained by random effects-only model (Items and Speakers)	$R2_{cond} = 27\%$ $C = 0.773$	$R2_{cond} = 39\%$ $C = 0.835$	$R2_{cond} = 68\%$ $C = 0.949$
Variance explained by fixed effects-only model	$C = 0.752$	$C = 0.697$	$C = 0.743$

Note: Cells with ‘NA’ indicate interaction effects that could not be included in the respective model due to data sparsity. Cells with ‘NS’ signal that the interaction effect noted in one of the other models are not significant in the respective model.

* $p < .05$; ** $p < .01$; *** $p < .001$.